

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:18:11 ; Search time 22.1667 Seconds
(without alignments)
75.512 Million cell updates/sec

Title: US-09-870-089B-13

Perfect score: 55

Sequence: 1 KHFKPHGFS 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	55	100.0	9	12	US-09-870-089B-13
2	55	100.0	286	10	US-09-925-300-1296
3	55	100.0	351	12	US-09-870-089B-2
4	55	100.0	351	12	US-10-341-434-67
5	43	78.2	348	12	US-09-882-227-252
6	37	67.3	17	12	US-10-116-788A-8
7	37	67.3	110	12	US-10-116-788A-10
8	35	63.6	96	9	US-09-864-761-47039
9	35	63.6	244	9	US-09-815-242-10094
10	35	63.6	244	9	US-09-815-242-10094
11	35	63.6	384	12	US-10-120-801-89
12	35	63.6	512	10	US-09-983-204-14
13	35	63.6	512	12	US-10-258-073-2
14	35	63.6	512	12	US-10-258-073-6
15	35	63.6	710	12	US-10-032-585-7813

16	35	63.6	908	9	US-09-895-072-15	Sequence 15, Appl
17	35	63.6	908	10	US-09-986-552-15	Sequence 15, Appl
18	35	63.6	908	12	US-10-023-894-9	Sequence 9, Appli
19	35	63.6	908	12	US-10-024-197-9	Sequence 9, Appli
20	35	63.6	908	12	US-10-306-686-15	Sequence 15, Appl
21	35	63.6	908	15	US-10-023-888-9	Sequence 9, Appli
22	35	63.6	908	15	US-10-023-889-9	Sequence 9, Appli
23	35	63.6	908	15	US-10-023-890-9	Sequence 9, Appli
24	34	61.8	50	9	US-09-815-242-5741	Sequence 5741, Ap
25	34	61.8	50	9	US-09-815-242-12702	Sequence 12702, A
26	34	61.8	50	9	US-09-815-242-12812	Sequence 12812, A
27	34	61.8	121	12	US-10-291-851-47	Sequence 47, Appl
28	34	61.8	215	15	US-10-156-761-8043	Sequence 8043, Ap
29	34	61.8	395	12	US-10-017-161-2154	Sequence 2154, Ap
30	34	61.8	405	10	US-09-854-731-20	Sequence 20, Appl
31	33	60.0	223	15	US-10-156-761-8301	Sequence 8301, Ap
32	33	60.0	247	9	US-09-815-242-5419	Sequence 5419, Ap
33	33	60.0	250	9	US-09-815-242-12563	Sequence 12563, A
34	33	60.0	250	11	US-09-966-521-32	Sequence 32, Appl
35	33	60.0	250	12	US-10-429-094-32	Sequence 32, Appl
36	33	60.0	272	15	US-10-102-627-62	Sequence 62, Appl
37	33	60.0	299	11	US-09-934-455-318	Sequence 318, App
38	33	60.0	299	16	US-10-278-536-64	Sequence 64, Appl
39	33	60.0	382	15	US-10-100-121-18	Sequence 18, Appl
40	33	60.0	722	15	US-10-043-487-263	Sequence 263, App
41	33	60.0	731	10	US-09-738-626-4854	Sequence 4854, Ap
42	33	60.0	786	15	US-10-045-792-11	Sequence 11, Appl
43	33	60.0	791	12	US-10-320-800-16	Sequence 16, Appl
44	33	60.0	796	12	US-10-276-629-16	Sequence 16, Appl
45	33	60.0	809	12	US-10-032-585-7691	Sequence 7691, Ap

ALIGNMENTS

RESULT 1

US-09-870-089B-13
; Sequence 13, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ATF4/CREB-2
US-09-870-089B-13

Query Match 100.0%; Score 55; DB 12; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
DB 1 KHFKPHGFS 9

RESULT 2

US-09-925-300-1296
; Sequence 1296, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101

CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1296
LENGTH: 286
TYPE: PRT
ORGANISM: Homo sapiens
S-09-925-300-1296

Query Match 100.0%; Score 55; DB 10; Length 286;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
DB 53 KHFKPHGFS 61

RESULT 3

US-09-870-089B-2
; Sequence 2, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-870-089B-2

Query Match 100.0%; Score 55; DB 12; Length 351;
Best Local Similarity 100.0%; Pred. No. 0.074;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
DB 42 KHFKPHGFS 50

RESULT 4

US-10-341-434-67
; Sequence 67, Application US/10341434
; Publication No. US20030215935A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-67

Query Match 100.0%; Score 55; DB 12; Length 351;

Best Local Similarity 100.0%; Pred. No. 0.074;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
DB 42 KHFKPHGFS 50

RESULT 5

US-09-882-227-252
; Sequence 252, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 252
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-252

Query Match 78.2%; Score 43; DB 12; Length 348;
Best Local Similarity 87.5%; Pred. No. 9.6;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 HFKPHGFS 9
DB 16 HFKPHDFS 23

RESULT 6

US-10-116-788A-8
; Sequence 8, Application US/10116788A
; Publication No. US20030170840A1
; GENERAL INFORMATION:
; APPLICANT: McDonald, Thomas L.
; APPLICANT: Larson, Marilyn A.
; APPLICANT: Weber, Annika
; TITLE OF INVENTION: Genomic Mammary Amyloid A Sequence
; FILE REFERENCE: P04557US1
; CURRENT APPLICATION NUMBER: US/10/116,788A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 09/425,679
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/218,482
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/218,611
; PRIOR FILING DATE: 2000-07-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Equus caballus
US-10-116-788A-8

Query Match 67.3%; Score 37; DB 12; Length 17;
Best Local Similarity 83.3%; Pred. No. 5;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 HFPHG 7
 ||:||||
 Db 7 HFRPHG 12

RESULT 7
 US-10-116-788A-10
 ; Sequence 10, Application US/10116788A
 ; Publication No. US20030170840A1
 ; GENERAL INFORMATION:
 ; APPLICANT: McDonald, Thomas L.
 ; APPLICANT: Larson, Marilyn A.
 ; APPLICANT: Weber, Annika
 ; TITLE OF INVENTION: Genomic Mammary Amyloid A Sequence
 ; FILE REFERENCE: P04557US1
 ; CURRENT APPLICATION NUMBER: US/10/116,788A
 ; CURRENT FILING DATE: 2002-06-10
 ; PRIOR APPLICATION NUMBER: 09/425,679
 ; PRIOR FILING DATE: 1999-10-22
 ; PRIOR APPLICATION NUMBER: 60/218,482
 ; PRIOR FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: 60/218,611
 ; PRIOR FILING DATE: 2000-07-17
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 10
 ; LENGTH: 110
 ; TYPE: PRT
 ; ORGANISM: Equus caballus
 US-10-116-788A-10

Query Match 67.3%; Score 37; DB 12; Length 110;
 Best Local Similarity 83.3%; Pred. No. 34;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 HFPHG 7
 ||:||||
 Db 100 HFRPHG 105

RESULT 8
 US-09-864-761-47039
 ; Sequence 47039, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecomica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 47039
 ; LENGTH: 96
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC004147.1
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.71
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.52
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.58
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
 ; OTHER INFORMATION: EST HUMAN HIT: A153866.1, EVALUATE 5.00e-25
 ; OTHER INFORMATION: SWISSPROT HIT: Q16515, EVALUATE 2.00e-51
 US-09-864-761-47039

Query Match 63.6%; Score 35; DB 9; Length 96;
 Best Local Similarity 66.7%; Pred. No. 66;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
 ||:||||
 Db 50 KHYKPKQFS 58

RESULT 9
 US-09-815-242-10094
 ; Sequence 10094, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011a
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308

IOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 ID NO 10094
 LENGTH: 244
 TYPE: PRT
 ORGANISM: Escherichia coli
 9-815-242-10094

Query Match 63.6%; Score 35; DB 9; Length 244;
 st Local Similarity 71.4%; Pred. No. 1.7e+02;
 tches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 KHFKPHG 7
 : |||||
 101 RHVKPHG 107

JUL 10
 09-815-242-13891
 sequence 13891, Application US/09815242
 atent No. US2002061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Orlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
 Prokaryotes

FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13891
 LENGTH: 244
 TYPE: PRT
 ORGANISM: Salmonella typhi
 US-09-815-242-13891

Query Match 63.6%; Score 35; DB 9; Length 244;
 Best Local Similarity 71.4%; Pred. No. 1.7e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHFKPHG 7
 : |||||
 DB 101 RHVKPHG 107

RESULT 11
 US-10-120-801-89
 Sequence 89, Application US/10120801
 Publication No. US20030203843A1
 GENERAL INFORMATION:
 APPLICANT: Fena, Carol

APPLICANT: Guo, Xiaojia
 APPLICANT: Shimkets, Richard
 APPLICANT: Padigaru, Muralidhara
 APPLICANT: Kekuda, Ramesh
 APPLICANT: Spytek, Kimberly
 APPLICANT: Menraban, Fuad N.
 APPLICANT: Topper, James N.
 APPLICANT: Wasserman, Scott
 APPLICANT: Edinger, Shlomit
 APPLICANT: Smithson, Glenna
 APPLICANT: Gunther, Erik
 APPLICANT: Komuves, Laszlo

TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-340
 CURRENT APPLICATION NUMBER: US/10/120,801
 CURRENT FILING DATE: 2002-04-11
 PRIOR APPLICATION NUMBER: 60/285748
 PRIOR FILING DATE: 2001-04-23
 PRIOR APPLICATION NUMBER: 60/286068
 PRIOR FILING DATE: 2001-04-24
 PRIOR APPLICATION NUMBER: 60/286292
 PRIOR FILING DATE: 2001-04-25
 PRIOR APPLICATION NUMBER: 60/288334
 PRIOR FILING DATE: 2001-05-03
 PRIOR APPLICATION NUMBER: 60/291241
 PRIOR FILING DATE: 2001-05-16
 PRIOR APPLICATION NUMBER: 60/322284
 PRIOR FILING DATE: 2001-09-14
 PRIOR APPLICATION NUMBER: 60/285609
 PRIOR FILING DATE: 2001-04-20
 NUMBER OF SEQ ID NOS: 155
 SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 89
 LENGTH: 384
 TYPE: PRT
 ORGANISM: mouse
 US-10-120-801-89

Query Match 63.6%; Score 35; DB 12; Length 384;
 Best Local Similarity 62.5%; Pred. No. 2.7e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 HFKPHGFS 9
 : |||||
 DB 129 HRRPHGFA 136

RESULT 12
 US-09-983-204-14
 Sequence 14, Application US/09983204
 Patent No. US20020173000A1
 GENERAL INFORMATION:
 APPLICANT: RENARD, STEPHANE
 APPLICANT: BESNARD, FRANCOIS
 APPLICANT: GRAHAM, DAVID
 TITLE OF INVENTION: SODIUM CHANNEL RECEPTOR
 FILE REFERENCE: 07586.0010
 CURRENT APPLICATION NUMBER: US/09/983,204
 CURRENT FILING DATE: 2001-10-23
 PRIOR APPLICATION NUMBER: 09/424,666
 PRIOR FILING DATE: 2001-02-22
 PRIOR APPLICATION NUMBER: PCT/EP98/02884
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 97401196.7
 PRIOR FILING DATE: 1997-05-30
 NUMBER OF SEQ ID NOS: 19
 SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14
 LENGTH: 512
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:

OTHER INFORMATION: MDEG
US-09-983-204-14

Query Match 63.6%; Score 35; DB 10; Length 512;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
||:||||
Db 144 KHYKPKQFS 152

RESULT 13

US-10-258-073-2
; Sequence 2, Application US/10258073
; Publication No. US20030219858A1
; GENERAL INFORMATION:
; APPLICANT: McGill University
; APPLICANT: Babinaki, Kazimierz
; APPLICANT: Seguela, Philippe
; TITLE OF INVENTION: A NOVEL HETEROMULTIMERIC ION CHANNEL RECEPTOR AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 0103.001-WO-US
; CURRENT APPLICATION NUMBER: US/10/258,073
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: PCT/CA01/00561
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 512
; TYPE: PRT
; ORGANISM: HUMAN ASIC2A
US-10-258-073-2

Query Match 63.6%; Score 35; DB 12; Length 512;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
||:||||
Db 144 KHYKPKQFS 152

RESULT 14

US-10-258-073-6
; Sequence 6, Application US/10258073
; Publication No. US20030219858A1
; GENERAL INFORMATION:
; APPLICANT: McGill University
; APPLICANT: Babinaki, Kazimierz
; APPLICANT: Seguela, Philippe
; TITLE OF INVENTION: A NOVEL HETEROMULTIMERIC ION CHANNEL RECEPTOR AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 0103.001-WO-US
; CURRENT APPLICATION NUMBER: US/10/258,073
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: PCT/CA01/00561
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 512
; TYPE: PRT
; ORGANISM: RAT ASIC2A
US-10-258-073-6

Query Match 63.6%; Score 35; DB 12; Length 512;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
||:||||

Db 144 KHYKPKQFS 152

RESULT 15

US-10-032-585-7813
; Sequence 7813, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: BO, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7813
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7813

Query Match 63.6%; Score 35; DB 12; Length 710;
Best Local Similarity 75.0%; Pred. No. 5.1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KHFKPHGF 8
|||||
Db 572 KHFKKHEF 579

Search completed: December 16, 2003, 14:41:07
Job time : 22.1667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:05 ; Search time 33 Seconds
(without alignments)
43.289 Million cell updates/sec

Title: US-09-870-089B-13

Perfect score: 55

Sequence: 1 KHFKPHGFS 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	100.0	9	AAU74686	Human cancer antigen
2	55	100.0	93	ABG17297	Novel human antigen
3	55	100.0	99	AAAG00153	Human secreted protein
4	55	100.0	121	ABG17298	Novel human antigen
5	55	100.0	286	AA856718	Human prostate cancer
6	55	100.0	351	AAU74680	Human cancer antigen
7	55	100.0	366	ABG17300	Novel human antigen
8	55	100.0	410	ABG17299	Novel human antigen
9	49	89.1	346	ABB57035	Mouse ischaemic CO

10	43	78.2	348	19	AAW98693	H. pylori GHP0 632
11	41	74.5	643	22	ABG24303	Novel human antigen
12	41	74.5	2391	15	AAW55694	Carbamoyl-phosphat
13	40	72.7	119	19	AAW56455	Protein encoded by
14	38	69.1	1369	22	ABG60839	Drosophila melanog
15	37	67.3	17	22	AAE01067	Horse colostrum-as
16	37	67.3	110	22	AAE01069	Horse serum SAA (s
17	37	67.3	1288	13	AAW20073	N-methylhydantoina
18	36	65.5	366	24	ABP77778	N. gonorrhoeae ami
19	36	65.5	374	22	ABG68379	Drosophila melanog
20	36	65.5	1325	24	ABU56657	Lung cancer-associ
21	35	63.6	95	24	ABG99937	Human novel polype
22	35	63.6	96	22	ABG54134	Human liver peptid
23	35	63.6	96	22	AAW72503	Human bone marrow
24	35	63.6	96	22	AAW32749	Peptide #6786 enco
25	35	63.6	96	23	ABG42323	Human peptide enco
26	35	63.6	143	21	AAW24472	Arabidopsis thalia
27	35	63.6	174	21	AAW24711	Arabidopsis thalia
28	35	63.6	175	21	AAW24471	Arabidopsis thalia
29	35	63.6	175	24	ABP78104	Arabidopsis thalia
30	35	63.6	200	22	ABG13915	Novel human antigen
31	35	63.6	202	21	AAW24470	Arabidopsis thalia
32	35	63.6	206	21	AAW24710	Arabidopsis thalia
33	35	63.6	228	21	AAW24709	Arabidopsis thalia
34	35	63.6	244	22	AAU34501	Arabidopsis thalia
35	35	63.6	244	22	AAU38298	E. coli cellular p
36	35	63.6	250	22	ABG29092	Salmonella typhi c
37	35	63.6	258	21	AAW25505	Novel human antigen
38	35	63.6	258	21	AAW25562	Arabidopsis thalia
39	35	63.6	309	21	AAW25504	Arabidopsis thalia
40	35	63.6	309	21	AAW25561	Arabidopsis thalia
41	35	63.6	317	21	AAW25560	Arabidopsis thalia
42	35	63.6	317	21	AAW25560	Arabidopsis thalia
43	35	63.6	318	21	AAW34951	Arabidopsis thalia
44	35	63.6	330	22	ABG08518	Novel human antigen
45	35	63.6	350	21	AAW34950	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAU74686 standard; Peptide; 9 AA.

AAU74686;

09-APR-2002 (first entry)

Human cancer antigen ATF4/CREB-2 natural immunogenic ligand.

Human; cancer antigen; ATF4; CREB-2; vaccine; cytostatic;
immunogenic ligand; gene therapy; MHC; major histocompatibility complex;
adoptive immunotherapy; cancer; ovarian cancer.

Homo sapiens.

WO200192306-A2.

06-DEC-2001.

30-MAY-2001; 2001WO-US17454.

31-MAY-2000; 2000US-209388P.

20-DEC-2000; 2000US-257007P.

(GENZ) GENZYME CORP.

Nicolette CA;

WPI; 2002-097764/13.

New therapeutic compounds comprising immunogenic ligands, useful for

PT modulating an immune response, particularly for treating ovarian
 XX cancer, and as components of anticancer vaccines -
 XX
 PS Claim 6; Page 56; 68pp; English.
 XX
 CC The invention relates to compounds comprising an immunogenic ligand
 CC whose sequence is based in part on residues 42-50 of human cancer
 CC antigen Atp4/CREB-2 (not defined) and the polynucleotides encoding them.
 CC Also included are an antibody that specifically recognises and binds the
 CC compound, a method for inducing an immune response in a subject by
 CC delivering the compound, a method of immunotherapy comprising
 CC administering to a subject the antibody, an immune effector cell that has
 CC been raised in vitro or in vivo in the presence and at the expense of an
 CC antigen presenting cell that presents the immunogenic compound in the
 CC context of an MHC (major histocompatibility complex) molecule and a
 CC method of adoptive immunotherapy comprising administering the immune
 CC effector cell. The compounds are useful for modulating an immune response
 CC to the synthetic and naturally occurring compounds. The compounds are
 CC especially useful in gene therapy or as components of anti-cancer
 CC vaccines. The compounds are useful for treating cancer, particularly
 CC ovarian cancer. The compounds are also useful for generating antibodies
 CC that specifically recognise and bind to these molecules. These
 CC antibodies are further useful for immunotherapy when administered to a
 CC subject. The peptides, polypeptides and polynucleotides are useful in
 CC diagnostic methods, for the detection and purification of antibodies,
 CC or as immunogens for the production of antibodies. The present
 CC sequence represents a human cancer antigen Atp4/CREB-2 based immunogenic
 CC ligand of the invention.
 CC Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
 CC encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
 CC but have not been cross-referenced or CDS features put in due to the
 CC degeneracy of the DNA sequences.

XX
 SQ Sequence 9 AA;

Query Match 100.0%; Score 55; DB 23; Length 9;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
 |||||
 Db 1 KHFKPHGFS 9

RESULT 2
 ABG17297
 ID ABG17297 standard; Protein; 93 AA.

XX
 AC ABG17297;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #17288.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS81484.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PT

XX Claim 20; SEQ ID No 47656; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 93 AA;

Query Match 100.0%; Score 55; DB 22; Length 93;
 Best Local Similarity 100.0%; Pred. No. 0.0099;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
 |||||
 Db 38 KHFKPHGFS 46

RESULT 3
 AAG00153
 ID AAG00153 standard; Protein; 99 AA.

XX
 AC AAG00153;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 4224.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

OS Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GSEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR N-PSDB; AAC00159.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 13; SEQ ID 4234; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 99 AA;

Query Match 100.0%; Score 55; DB 21; Length 99;
 Best Local Similarity 100.0%; Pred. NO. 0.011;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKPHGFS 9
 Db 74 KHKPHGFS 82

RESULT 4
 ABG17298

ID ABG17298 standard; Protein; 121 AA.

XX AC ABG17298;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #17289.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS81485.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX PS Claim 20; SEQ ID No 47657; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 121 AA;

Query Match 100.0%; Score 55; DB 22; Length 121;
 Best Local Similarity 100.0%; Pred. NO. 0.013;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKPHGFS 9
 Db 20 KHKPHGFS 28

RESULT 5

AAB56718

ID AAB56718 standard; Protein; 286 AA.

XX AC AAB56718;

XX DT 13-MAR-2001 (first entry)

XX DE Human prostate cancer antigen protein sequence SEQ ID NO:1296.

XX KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease.

XX OS Homo sapiens.

XX PN WO200055174-A1.

XX PD 21-SEP-2000.

XX XX 08-MAR-2000; 2000WO-US05988.

XX XX 12-MAR-1999; 99US-0124270.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX XX (ROSE/) ROSEN C A.

XX XX Rosen CA, Ruben SM;

XX XX WPI; 2000-587513/55.

XX XX N-PSDB; AAF15921.

XX PT Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -

XX PS Claim 11; Page 1716-1718; 2338pp; English.

XX CC AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,

CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
XX
SQ Sequence 286 AA;

Query Match 100.0%; Score 55; DB 21; Length 286;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
Db 53 KHFKPHGFS 61
|||||

RESULT 6
AAU74680
ID AAU74680 standard; Protein; 351 AA.
XX
AC AAU74680;
XX
DT 09-APR-2002 (first entry)
XX
DE Human cancer antigen ATP4/CREB-2.
XX
KW Human; cancer antigen; ATP4; CREB-2; vaccine; cytostatic;
KW immunogenic ligand; gene therapy; MHC; major histocompatibility complex;
KW adoptive immunotherapy; cancer; ovarian cancer.
XX
OS Homo sapiens.
XX
FN WO200192306-A2.
XX
PD 06-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US17454.
XX
PR 31-MAY-2000; 2000US-209388P.
PR 20-DEC-2000; 2000US-257007P.
XX
PA (GENZ) GENZYME CORP.
XX
PI Nicolette CA;
XX
DR WPI; 2002-097764/13.
DR N-PSDB; AAG20119.
XX
PT New therapeutic compounds comprising immunogenic ligands, useful for
PT modulating an immune response, particularly for treating ovarian
PT cancer, and as components of anticancer vaccines -
XX
PS Claim 7; Page 64; 68pp; English.
XX
CC The invention relates to compounds comprising an immunogenic ligand
CC whose sequence is based in part on residues 42-50 of human cancer
CC antigen ATP4/CREB-2 (not defined) and the polynucleotides encoding them.
CC Also included are an antibody that specifically recognises and binds the
CC compound, a method for inducing an immune response in a subject by
CC delivering the compound, a method of immunotherapy comprising
CC administering to a subject the antibody, an immune effector cell that has
CC been raised in vitro or in vivo in the presence and at the expense of an
CC antigen presenting cell that presents the immunogenic compound in the
CC context of an MHC (major histocompatibility complex) molecule and a
CC method of adoptive immunotherapy comprising administering the immune
CC effector cell. The compounds are useful for modulating an immune response
CC to the synthetic and naturally occurring compounds. The compounds are

CC especially useful in gene therapy or as components of anti-cancer
CC vaccines. The compounds are useful for treating cancer, particularly
CC ovarian cancer. The compounds are also useful for generating antibodies
CC that specifically recognise and bind to these molecules. These
CC antibodies are further useful for immunotherapy when administered to a
CC subject. The peptides, polypeptides and polynucleotides are useful in
CC diagnostic methods, for the detection and purification of antibodies,
CC or as immunogens for the production of antibodies. The present
CC sequence represents human cancer antigen ATP4/CREB-2 upon which the
CC immunogenic ligands of the invention are based.
XX
XX
SQ Sequence 351 AA;

Query Match 100.0%; Score 55; DB 23; Length 351;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
Db 42 KHFKPHGFS 50
|||||

RESULT 7
ABG17300
ID ABG17300 standard; Protein; 366 AA.
XX
AC ABG17300;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17291.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAG81487.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 47659; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 366 AA;

Query Match 100.0%; Score 55; DB 22; Length 366;
 Best Local Similarity 100.0%; Pred. No. 0.045;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
 | | | | | | | | |
 Db 41 KHFKPHGFS 49

RESULT 8
 ABG17299
 ID ABG17299 standard; Protein; 410 AA.

XX AC ABG17299;
 XX DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #17290.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS81486.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX PS Claim 20; SEQ ID NO 47658; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 410 AA;

Query Match 100.0%; Score 55; DB 22; Length 410;
 Best Local Similarity 100.0%; Pred. No. 0.051;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
 | | | | | | | | |
 Db 315 KHFKPHGFS 323

RESULT 9
 ABB57035
 ID ABB57035 standard; Protein; 346 AA.

XX AC ABB57035;

XX DT 07-MAR-2002 (first entry)

XX DE Mouse ischaemic condition related protein sequence SEQ ID NO:43.

XX KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease.

XX OS Mus musculus.

XX PN WO200188188-A2.

XX PD 22-NOV-2001.

XX PF 18-MAY-2001; 2001WO-JP04192.

XX PR 18-MAY-2000; 2000JP-0145977.

XX PA (UUNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

XX DR WPI; 2002-034733/04.

XX DR N-PSDB; ABI99228.

XX PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or
 PT by determining the expression profile of a gene group comprising these
 PT genes

XX PS Claim 2; Page 149-150; 2690pp; English.

XX The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (I) in a test sample or determining the expression profile of a
 CC gene group in the sample comprising genes selected from (I). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
 CC expression levels of particular genes (ABI99202 to ABI9912, encoding
 CC the protein sequences in ABB57020 to ABB57374) or by determining the
 CC expression profile of a gene group comprising these genes. The
 CC expression levels or expression profiles produced by these genes are
 CC used as an indicator when screening for ischaemic condition-improving
 CC drugs or therapeutics for ischaemic diseases. ABI9913 and ABI9914
 CC represent PCR primers for a mouse ischaemic condition related sequence,
 CC which are used in the exemplification of the present invention.

XX SQ Sequence 346 AA;

Query Match 89.1%; Score 49; DB 23; Length 346;

Best Local Similarity 88.9%; Pred. No. 0.55;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HKFKPHGFS 9
Db 39 KHLKPHGFS 47

RESULT 10
AAW98693
ID AAW98693 standard; Protein; 348 AA.
XX
AC AAW98693;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 632 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; AAX14412.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
XX Claim 8; Page 1444-1446; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 348 AA;

Query Match 78.2%; Score 43; DB 19; Length 348;
Best Local Similarity 87.5%; Pred. No. 7.2;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 HKFKPHGFS 9
Db 16 HKFPHDFS 23

RESULT 11
ABG24303
ID ABG24303 standard; Protein; 643 AA.
XX
AC ABG24303;
XX
DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #24294.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS88490.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 54662; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 643 AA;

Query Match 74.5%; Score 41; DB 22; Length 643;
Best Local Similarity 87.5%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 HKFKPHGFS 9
Db 67 HKFPHGQS 74

RESULT 12
AAR55694
ID AAR55694 standard; Protein; 2391 AA.
XX
AC AAR55694;
XX
DT 25-MAR-2003 (updated)
DT 06-DEC-1994 (first entry)
XX

DE Carbamoyl-phosphate-synthetase II.
 KW Carbamoyl-phosphate-synthetase II; CPSII; psCPSII gene; malaria.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 FT Domain 1..690 /note= "glutamine-amidotransferase domain"
 FT Domain 1..270 /note= "structural subdomain"
 FT Peptide 271..482 /note= "insert sequence"
 FT Domain 483..690 /note= "glutaminase subdomain"
 FT Domain 691..2391 /note= "Carbamoyl-phosphate-synthase domain"
 FT Domain 691..1254 /note= "ATP binding subdomain CP8a"
 FT Peptide 1255..1857 /note= "insert sequence"
 FT Domain 1858..2391 /note= "ATP binding subdomain CP8b"
 XX
 PN WO9412643-A1.
 XX
 XX 09-JUN-1994.
 XX
 XX 02-DEC-1993; 93WO-AU00617.
 XX
 XX 03-DEC-1992; 92AU-0006206.
 XX 16-DEC-1992; 92AU-0006380.
 XX (UNIX) UNISEARCH LTD.
 XX Flores MV, Osullivan WJ, Stewart TS;
 XX WPI; 1994-200271/24.
 XX N-PSDB; AAQ62924.
 XX
 XX Nucleic acid encoding carbamoyl phosphate synthetase II - isolated from Plasmodium falciparum, used to develop prods. for the treatment of malaria.
 PT
 PT
 PT
 XX Disclosure; Page 6-16; 31pp; English.
 XX
 CC The cDNA sequence encoding the carbamoyl-phosphate-transferase II (CPSII) of Plasmodium falciparum was determined. The cDNA encodes a protein that includes 2 insert sequences not found in other CPSII proteins. The first separates the putative structural subdomain and the glutaminase subdomain of the glutamine-amidotransferase subunit of CPSII, while the second separates 2 ATP binding subdomains of the CPSII subunit, CP8a and CP8b.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 2391 AA;
 Query Match 74.5%; Score 41; DB 15; Length 2391;
 Best Local Similarity 77.8%; Pred. No. 1.4e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KHFKPHGFS 9
 DB 1168 KYFKKHGFS 1176
 RESULT 13
 AAW56455
 ID AAW56455 standard; Protein; 119 AA.
 XX
 AC AAW56455;
 XX

DT 05-AUG-1998 (first entry)
 XX Protein encoded by testis 5' RACE clone type 1.
 XX
 KW Human; cancer cachectic factor; p24K-inc; weight loss; treatment; mouse tumour derived proteoglycan; cachectic effect; prevention; non-insulin dependent diabetes; NIDDM; syndrome X; diagnosis.
 KW
 XX Homo sapiens.
 OS
 XX WO9811136-A1.
 PN
 XX 19-MAR-1998.
 PD
 XX 09-SEP-1997; 97WO-DK00377.
 PF
 XX 09-SEP-1996; 96DK-0000968.
 PR
 XX (NOVO) NOVO-NORDISK AS.
 PA
 XX Clausen J, Din N, Farrah TM, Rasmussen PB, Vissing H;
 PI WPI; 1998-207329/18.
 XX N-PSDB; AAV23110.
 DR
 XX DNA encoding a cancer cachectic factor - used to prevent or treat non-insulin dependent diabetes or Syndrome X
 PT
 PT
 XX Example 1; Page 36; 56pp; English.
 PS
 XX
 CC The present sequence is encoded by a testis 5' RACE clone type 1, created to characterise the 5' end of the human cancer cachectic factor (CCF) open reading frame. The coding region of CCF has extensive homology to a 20 amino acid sequence obtained from the peptide core of a mouse tumour derived proteoglycan with cachectic effects. CCF, a 24 kD protein factor, induces weight loss, and can be used with a substance capable of reducing appetite or inducing satiety. CCF can alternatively be administered with a substance capable of increasing the ratio of lean to fat body mass. CCF is useful for treating conditions or disorders arising from obesity or for inducing weight loss. CCF is used to prevent or treat non-insulin dependent diabetes (NIDDM) or Syndrome X. Anti-CCF antibodies can be used for direct application to cancer patients in order to reduce weight loss and tumour growth. The antibodies can also be used to diagnose cancer by detecting the presence of p24CCF in urine or serum of potential cancer patients. Antisense CCF gene constructs are useful for the prevention of unwanted endogenous synthesis or activity of CCF, e.g. in potential cancer patients, thus preventing weight loss (tissue degeneration). Labelled CCF or pre CCF can be used to identify CCF receptors which are then useful for detection of CCF24 kD CxP core peptide agonists.
 CC
 SQ Sequence 119 AA;
 Query Match 72.7%; Score 40; DB 19; Length 119;
 Best Local Similarity 77.8%; Pred. No. 8;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KHFKPHGFS 9
 DB 7 KHFKPHSRS 15
 RESULT 14
 ABB60839
 ID ABB60839 standard; Protein; 1369 AA.
 XX
 AC ABB60839;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster polypeptide SEQ ID NO 9309.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW

KW pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX N-PSDB; ABL04942.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Disclosure; SEQ ID NO 9309; 2lpp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL6175) and the encoded proteins
 CC (ABB57737-ABB2072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1369 AA;

 Query Match 69.1%; Score 38; DB 22; Length 1369;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 3 FXPHGF 8
 DB 571 FXPHGF 576

 RESULT 15
 AAE01067
 ID AAE01067 standard; peptide; 17 AA.
 XX
 AC AAE01067;
 XX
 DT 17-JUL-2001 (first entry)
 XX
 DE Horse colostrum-associated serum amyloid A (SAA) peptide #5.
 XX
 KW Horse; colostrum; serum amyloid A; SAA; antibacterial; antidiarrhoeic;
 KW apolipoprotein; inflammatory response; mucin 3; MUC3; therapy;
 KW Escherichia coli infection; traveller's diarrhoea; infant diarrhoea;
 KW necrotising enterocolitis; NEC; urinary tract infection; dysentery;
 KW infectious diarrhoea.
 XX
 OS Equus caballus.
 OS
 PN WO200131006-A2.
 XX
 PD 03-MAY-2001.
 XX
 PF 20-OCT-2000; 2000WO-US29065.
 XX

PR 22-OCT-1999; 99US-0425679.
 PR 14-JUL-2000; 2000US-0218482.
 PR 17-JUL-2000; 2000US-0218611.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 XX
 PI McDonald TL, Weber A, Mack DR, Larson MA;
 XX WPI; 2001-308640/32.
 DR
 XX New colostrum-associated serum amyloid A protein induces mucin
 PT production in epithelial cells and is useful to prevent and treat
 PT enteric pathogen infection -
 XX
 XX Claim 3; Page 20; 8lpp; English.
 PS
 XX The patent discloses novel colostrum-associated Serum Amyloid A
 CC (SAA) from mammals. SAAs are small apolipoproteins that accumulate
 CC and associate rapidly with high-density lipoprotein 3 (HDL3) during
 CC the acute phase of inflammatory response. They are characterised
 CC by the TFLK motif in the N-terminal region which has the ability
 CC to stimulate mucin 3 (MUC3) production. Colostrum associated SAAs
 CC are used to prevent or treat infectious diseases associated with
 CC enteric pathogens (particularly Escherichia coli) such as traveller's
 CC diarrhoea, infant diarrhoea, necrotising enterocolitis (NEC), urinary
 CC tract infections, and infectious diarrhoea in herd animals. They are
 CC used to prevent dysentery and other infectious diseases particularly
 CC for the military.
 CC The present sequence is colostrum-associated serum amyloid A (SAA)
 CC peptide from horse.
 XX
 SQ Sequence 17 AA;

 Query Match 67.3%; Score 37; DB 22; Length 17;
 Best Local Similarity 83.3%; Pred. No. 3.4;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 QY 2 HFKPHG 7
 DB 7 HFRPHG 12

 Search completed: December 16, 2003, 14:14:31
 Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:05 ; Search time 33 Seconds
(without alignments)
43.289 Million cell updates/sec

Title: US-09-870-089B-11

Perfect score: 50

Sequence: 1 FLHKVHLYV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	9	23 AAU74685	Human cancer antigen
2	41	82.0	9	23 AAU74682	Human cancer antigen
3	40	80.0	148	21 AAB53260	Human colon cancer
4	40	80.0	528	22 AAB93552	Human protein sequ
5	40	80.0	735	22 AAM79281	Human protein sequ
6	40	80.0	735	22 AAB73094	Human muskelin. H
7	40	80.0	735	22 AAB27226	Human EXMAD-4 SEQ
8	36	72.0	127	21 AAG24107	Arabidopsis thalia
9	36	72.0	321	21 AAG06848	Arabidopsis thalia

10	36	72.0	329	21	AAG24453	Arabidopsis thalia
11	36	72.0	378	21	AAG06847	Arabidopsis thalia
12	36	72.0	454	23	ABB48228	Listeria monocytog
13	36	72.0	979	22	ABB63728	Drosophila melanog
14	36	72.0	1084	22	ABB65956	Drosophila melanog
15	35	70.0	9	23	AAU74683	Human cancer antigen
16	35	70.0	171	22	AAG82248	S. epidermidis ope
17	35	70.0	190	22	ABP40155	Staphylococcus epi
18	35	70.0	191	22	ABB11608	Human secreted pro
19	35	70.0	213	23	ABP73980	Candida albicans e
20	35	70.0	320	22	AAM25634	Human protein sequ
21	35	70.0	366	22	AAE11971	Human novel G-prot
22	35	70.0	587	24	ABP72227	Human G-protein co
23	35	70.0	763	22	AAE11970	Human novel G-prot
24	35	70.0	777	21	AAE11970	Human sensory tran
25	35	70.0	777	24	ABP81751	Human G protein-co
26	35	70.0	841	22	AAE10372	Human taste recept
27	35	70.0	841	22	AAE11969	Human novel G-prot
28	35	70.0	841	23	ABB79833	Human taste cell s
29	35	70.0	841	23	ABB77319	Human G-protein co
30	35	70.0	841	24	ABP70449	Amino acid sequenc
31	35	70.0	1040	22	ABG20154	Novel human diagno
32	34	68.0	56	22	AAM87172	Human immune/haema
33	34	68.0	91	22	AAO3783	Human polypeptide
34	34	68.0	96	22	AAO11603	Human polypeptide
35	34	68.0	110	20	AAE60516	Human normal blad
36	34	68.0	113	22	ABB96319	Human testicular a
37	34	68.0	113	22	AAM95788	Human reproductive
38	34	68.0	117	22	AAM64044	Human brain expres
39	34	68.0	125	22	AAU35458	Haemophilus influe
40	34	68.0	131	22	AAU34865	E. coli cellular p
41	34	68.0	131	22	AAU36039	Klebsiella pneumon
42	34	68.0	131	22	AAU38265	Salmonella typhi c
43	34	68.0	162	22	AAU28065	Novel human secret
44	34	68.0	202	22	AAU27531	Human G-Protein Co
45	34	68.0	311	19	AAM59530	Human peroxisomal

ALIGNMENTS

RESULT 1
AAU74685

ID AAU74685 standard; Peptide; 9 AA.

AC AAU74685;

DT 09-APR-2002 (first entry)

DE Human cancer antigen ATF4/CREB-2 based immunogenic ligand #5.

XX Human; cancer antigen; ATF4; CREB-2; vaccine; cytostatic; immunogenic ligand; gene therapy; MHC; major histocompatibility complex; adoptive immunotherapy; cancer; ovarian cancer.

OS Homo sapiens.

OS Synthetic.

PN WO200192306-A2.

PD 06-DEC-2001.

PF 30-MAY-2001; 2001WO-US17454.

PR 31-MAY-2000; 2000US-209388P.

PR 20-DEC-2000; 2000US-257007P.

FA (GENZ) GENZYME CORP.

PI Nicolette CA;

DR WPI; 2002-097764/13.

XX

CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 XX gastrointestinal disorders and inflammation.

SQ Sequence 735 AA;

Query Match 80.0%; Score 40; DB 22; Length 735;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYL 8

|||||||

Db 594 LHKVHYL 600

RESULT 8

AAG24107

ID AAG24107 standard; Protein; 127 AA.

XX AC

AC AAG24107;

XX DT

DT 17-OCT-2000 (first entry)

XX DE

DE Arabidopsis thaliana protein fragment SEQ ID NO: 27656.

XX EX

EX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX OS

OS Arabidopsis thaliana.

XX PN

PN EP1033405-A2.

XX PD

PD 06-SEP-2000.

XX PF

PF 25-FEB-2000; 2000EP-0301439.

XX PR

PR 25-FEB-1999; 99US-0121825.

PR PR

PR 05-MAR-1999; 99US-0123180.

PR PR

PR 09-MAR-1999; 99US-0123548.

PR PR

PR 23-MAR-1999; 99US-0123788.

PR PR

PR 25-MAR-1999; 99US-0126264.

PR PR

PR 29-MAR-1999; 99US-0126785.

PR PR

PR 01-APR-1999; 99US-0127462.

PR PR

PR 06-APR-1999; 99US-0128234.

PR PR

PR 08-APR-1999; 99US-0128714.

PR PR

PR 16-APR-1999; 99US-0129845.

PR PR

PR 19-APR-1999; 99US-0130077.

PR PR

PR 21-APR-1999; 99US-0130510.

PR PR

PR 23-APR-1999; 99US-0130891.

PR PR

PR 28-APR-1999; 99US-0131445.

PR PR

PR 30-APR-1999; 99US-0132048.

PR PR

PR 04-MAY-1999; 99US-0132407.

PR PR

PR 05-MAY-1999; 99US-0132484.

PR PR

PR 06-MAY-1999; 99US-0132485.

PR PR

PR 06-MAY-1999; 99US-0132486.

PR PR

PR 07-MAY-1999; 99US-0132487.

PR PR

PR 11-MAY-1999; 99US-0132863.

PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 21-JUN-1999; 99US-0139763.
 PR 22-JUN-1999; 99US-0139817.
 PR 23-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 28-JUN-1999; 99US-0140695.
 PR 29-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0140991.
 PR 01-JUL-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 06-JUL-1999; 99US-0142055.
 PR 08-JUL-1999; 99US-0142390.
 PR 09-JUL-1999; 99US-0142803.
 PR 12-JUL-1999; 99US-0142920.
 PR 13-JUL-1999; 99US-0142977.
 PR 14-JUL-1999; 99US-0143542.
 PR 15-JUL-1999; 99US-0143624.
 PR 16-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 19-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145224.
 PR 27-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.

Matches	6;	Conservative	1;	Mismatches	1;	Indels	0;	Gaps
QY	1	FLHKVHYL	8					
Db	50	FLHKIHL	57					
RESULT 9								
AAG06848								
ID	AAG06848 standard; Protein; 321 AA.							
XX								
AC	AAG06848;							
DT	17-OCT-2000 (first entry)							
XX								
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 3774.							
XX								
KW	Protein identification; signal transduction pathway; metabolic pathway;							
KW	hybridisation assay; Genetic mapping; Gene expression control; promoter;							
KW	termination sequence.							
XX								
OS	Arabidopsis thaliana.							
XX								
PN	EP1033405-A2.							
XX								
PD	06-SEP-2000.							
XX								
PF	25-FEB-2000; 2000EP-0301439.							
XX								
PR	25-FEB-1999; 99US-0121825.							
PR	05-MAR-1999; 99US-0123180.							
PR	09-MAR-1999; 99US-0123548.							
PR	23-MAR-1999; 99US-0125788.							
PR	25-MAR-1999; 99US-0126264.							
PR	29-MAR-1999; 99US-0126785.							
PR	01-APR-1999; 99US-0127462.							
PR	06-APR-1999; 99US-0128234.							
PR	08-APR-1999; 99US-0128714.							
PR	16-APR-1999; 99US-0129845.							
PR	19-APR-1999; 99US-0130077.							
PR	21-APR-1999; 99US-0130449.							
PR	23-APR-1999; 99US-0130510.							
PR	23-APR-1999; 99US-0130891.							
PR	28-APR-1999; 99US-0131449.							
PR	30-APR-1999; 99US-0132048.							
PR	30-APR-1999; 99US-0132407.							
PR	04-MAY-1999; 99US-0132484.							
PR	05-MAY-1999; 99US-0132485.							
PR	06-MAY-1999; 99US-0132486.							
PR	07-MAY-1999; 99US-0132487.							
PR	11-MAY-1999; 99US-0132863.							
PR	14-MAY-1999; 99US-0134256.							
PR	14-MAY-1999; 99US-0134218.							
PR	14-MAY-1999; 99US-0134219.							
PR	14-MAY-1999; 99US-0134221.							
PR	14-MAY-1999; 99US-0134370.							
PR	18-MAY-1999; 99US-0134768.							
PR	19-MAY-1999; 99US-0134941.							
PR	20-MAY-1999; 99US-0135124.							
PR	21-MAY-1999; 99US-0135353.							
PR	24-MAY-1999; 99US-0135629.							
PR	25-MAY-1999; 99US-0136021.							
PR	27-MAY-1999; 99US-0136392.							
PR	28-MAY-1999; 99US-0136782.							
PR	01-JUN-1999; 99US-0137222.							
PR	03-JUN-1999; 99US-0137528.							
PR	04-JUN-1999; 99US-0137502.							
PR	07-JUN-1999; 99US-0137724.							
PR	08-JUN-1999; 99US-0138094.							
PR	10-JUN-1999; 99US-0138540.							
PR	10-JUN-1999; 99US-0138847.							
PR	14-JUN-1999; 99US-0139119.							
PR	16-JUN-1999; 99US-0139452.							

Query Match

Best Local Similarity

72.0%;

75.0%;

Score 36;

DB 21;

Length 127;

Pred. No. 33;

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 72.0%; Score 36; DB 21; Length 321;
Best Local Similarity 71.4%; Pred. No. 87;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHYLV 9
|||:|:|
Db 32 HKIHVVV 38


```
RESULT 10
AAG24453
ID AAG24453 standard; Protein; 329 AA.
XX AC AAG24453;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 28130.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; Genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137503.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138684.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
```

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 72.0%; Score 36; DB 21; Length 329;

Best Local Similarity 75.0%; Pred. No. 89;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHYL 8

Db 52 FLHKIHL 59

RESULT 11

AAG06847

ID AAG06847 standard; Protein; 378 AA.

XX AAG06847;

AC AAG06847;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 3773.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 08-APR-1999; 99US-0128234.
XX 16-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0128845.
XX 21-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 72.0%; Score 36; DB 21; Length 378;
Best Local Similarity 71.4%; Pred. NO. 1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Caps 0;

QY 3 HKVHVLV 9
||:|:|
Db 89 HKIHVV 95

RESULT 12

ABB48228
ID ABB48228 standard; Protein; 454 AA.
XX
AC ABB48228;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #932.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.

XX WO200177335-A2.
 XX PD 18-OCT-2001.
 XX PF 11-APR-2001; 2001WO-FR01118.
 XX PR 11-APR-2000; 2000FR-0004629.
 XX PA (INSP) INST PASTEUR.
 XX PI Buchrieser C, Fzangeul L, Couve E, Rusniok C, Psihi H, Deboux P;
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX PS WPI; 2002-010914/01.
 XX DR Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
 XX PT and prevention of *Listeria* and related bacterial infections, and
 XX PT related polypeptides -
 XX PS Claim 6; SEQ ID No 933; 192pp; French.
 XX CC The present invention relates to the genome sequence of *Listeria*
 CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in *L.*
 CC *monocytogenes* and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of *L. monocytogenes* and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by *L.*
 CC *monocytogenes* and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 454 AA;
 Query Match 72.0%; Score 36; DB 23; Length 454;
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 FLHKVHY 7
 :|||:
 Db 239 YLKLHY 245
 RESULT 13
 ABB63728
 ID ABB63728 standard; Protein; 979 AA.
 XX AC ABB63728;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 17976.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX

PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656960/75.
 XX DR N-PSDB; ABL07831.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -
 XX PS Disclosure; SEQ ID NO 17976; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 979 AA;
 Query Match 72.0%; Score 36; DB 22; Length 979;
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 FLHKVHY 7
 :|||:
 Db 663 FIHKVHF 669
 RESULT 14
 ABB65956
 ID ABB65956 standard; Protein; 1084 AA.
 XX AC ABB65956;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 24660.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX

DR WPI; 2001-656860/75.
 XX N-PSDB; ABL10059.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 24660; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1084 AA;
 Query Match 72.0%; Score 36; DB 22; Length 1084;
 Best Local Similarity 55.6%; Pred. No. 3.1e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHYLV 9
 Db 345 FLNHLHLI 353
 ||| :|:
 345 FLNHLHLI 353
 RESULT 15
 AAU74683
 ID AAU74683 standard; Peptide; 9 AA.
 AC AAU74683;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Human cancer antigen ATF4/CREB-2 based immunogenic ligand #3.
 XX
 KW Human; cancer antigen; ATF4; CREB-2; vaccine; cytostatic;
 KW immunogenic ligand; gene therapy; MHC; major histocompatibility complex;
 KW adoptive immunotherapy; cancer; ovarian cancer.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200192306-A2.
 XX
 PD 06-DEC-2001.
 XX
 XX 30-MAY-2001; 2001WO-US17454.
 XX
 XX 31-MAY-2000; 2000US-209388P.
 PR
 PR 20-DEC-2000; 2000US-257007P.
 XX
 PA (GENZ) GENZYME CORP.
 XX
 PI Nicolette CA;
 XX
 XX WPI; 2002-097764/13.
 DR
 XX
 PT New therapeutic compounds comprising immunogenic ligands, useful for
 PT modulating an immune response, particularly for treating ovarian
 PT cancer, and as components of anticancer vaccines -
 XX
 XX Claim 3; Page 55; 68pp; English.
 PS
 XX The invention relates to compounds comprising an immunogenic ligand
 CC whose sequence is based in part on residues 42-50 of human cancer
 CC antigen ATF4/CREB-2 (not defined) and the polynucleotides encoding them.

CC Also included are an antibody that specifically recognises and binds the
 CC compound, a method for inducing an immune response in a subject by
 CC delivering the compound, a method of immunotherapy comprising
 CC administering to a subject the antibody, an immune effector cell that has
 CC been raised in vitro or in vivo in the presence and at the expense of an
 CC antigen presenting cell that presents the immunogenic compound in the
 CC context of an MHC (major histocompatibility complex) molecule and a
 CC method of adoptive immunotherapy comprising administering the immune
 CC effector cell. The compounds are useful for modulating an immune response
 CC to the synthetic and naturally occurring compounds. The compounds are
 CC especially useful in gene therapy or as components of anti-cancer
 CC vaccines. The compounds are useful for treating cancer, particularly
 CC ovarian cancer. The compounds are also useful for generating antibodies
 CC that specifically recognise and bind to these molecules. These
 CC antibodies are further useful for immunotherapy when administered to a
 CC subject. The peptides, polypeptides and polynucleotides are useful in
 CC diagnostic methods, for the detection and purification of antibodies,
 CC or as immunogens for the production of antibodies. The present
 CC sequence represents a human cancer antigen ATF4/CREB-2 based immunogenic
 CC ligand of the invention.
 CC Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
 CC encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
 CC but have not been cross-referenced or CDS features put in due to the
 CC degeneracy of the DNA sequences.
 XX
 SQ Sequence 9 AA;
 Query Match 70.0%; Score 35; DB 23; Length 9;
 Best Local Similarity 66.7%; Pred. No. 9.3e+05;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHYLV 9
 Db 1 FLHKVHWV 9
 ||| :|:
 1 FLHKVHWV 9

Search completed: December 16, 2003, 14:14:30
 Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:11:00 ; Search time 12.3333 Seconds
(without alignments)
30.875 Million cell updates/sec

Title: US-09-870-089B-11

Perfect score: 50

Sequence: 1 FLHKVHVLV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	72.0	114	2	US-08-378-939-36
2	35	70.0	190	4	US-09-134-001C-5000
3	34	68.0	311	1	US-08-872-784-1
4	34	68.0	311	2	US-09-100-851-1
5	34	68.0	311	3	US-09-285-294-1
6	34	68.0	1068	4	US-09-215-694-7
7	33	66.0	144	4	US-09-328-352-7101
8	33	66.0	157	4	US-08-624-125-19
9	33	66.0	288	1	US-08-937-155-19
10	33	66.0	288	4	US-08-136-420A-20
11	33	66.0	471	4	US-09-499-302A-7
12	33	66.0	502	4	US-09-252-991A-23106
13	33	66.0	554	4	US-08-190-802A-63
14	33	66.0	713	1	US-08-477-346-63
15	33	66.0	713	3	US-08-473-089-63
16	33	66.0	713	4	US-08-487-072A-63
17	33	66.0	713	4	US-09-108-857-3
18	33	66.0	713	4	US-07-752-101A-6
19	32	64.0	25	1	US-07-752-101A-13
20	32	64.0	37	1	US-07-752-101A-13
21	32	64.0	122	4	US-09-732-210-1619
22	32	64.0	195	1	US-07-752-101A-68
23	32	64.0	222	4	US-09-252-991A-23106
24	32	64.0	311	4	US-09-328-352-5381
25	32	64.0	321	4	US-09-107-532A-4118
26	32	64.0	353	1	US-07-752-101A-35
27	32	64.0	353	1	US-07-752-101A-36

28	32	64.0	354	1	US-07-752-101A-38	Sequence 38, Appl
29	32	64.0	354	1	US-07-752-101A-39	Sequence 39, Appl
30	32	64.0	354	1	US-07-752-101A-41	Sequence 41, Appl
31	32	64.0	354	1	US-07-752-101A-51	Sequence 51, Appl
32	32	64.0	375	1	US-07-752-101A-52	Sequence 52, Appl
33	32	64.0	672	1	US-07-841-651-2	Sequence 2, Appl
34	32	64.0	672	1	US-07-841-651-3	Sequence 3, Appl
35	32	64.0	705	4	US-09-328-352-7436	Sequence 7436, Ap
36	32	64.0	2938	5	PCT-US94-00198-3	Sequence 3, Appl
37	31	62.0	142	4	US-09-252-991A-33130	Sequence 33130, A
38	31	62.0	235	4	US-09-328-352-4407	Sequence 4407, Ap
39	31	62.0	261	2	US-08-484-905-111	Sequence 111, App
40	31	62.0	261	3	US-08-481-985B-111	Sequence 111, App
41	31	62.0	261	3	US-08-370-476-111	Sequence 111, App
42	31	62.0	266	3	US-08-976-255-13	Sequence 13, Appl
43	31	62.0	341	4	US-09-739-455-7	Sequence 7, Appl
44	31	62.0	341	4	US-09-739-455-17	Sequence 17, Appl
45	31	62.0	390	4	US-09-134-001C-3112	Sequence 3112, Ap

ALIGNMENTS

RESULT 1
US-08-378-939-36
; Sequence 36, Application US/08378939
; Patent No. 5876961
; GENERAL INFORMATION:
; APPLICANT: CROME, JAMES SCOTT
; APPLICANT: LEWIS, ALAN PETER
; TITLE OF INVENTION: PRODUCTION OF ANTIBODIES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
; STREET: 555 THIRTEENTH ST. N.W.
; CITY: WASHINGTON
; STATE: D. C.
; COUNTRY: U.S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,939
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952640
; FILING DATE: 01-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-118
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-378-939-36

Query Match 72.0%; Score 36; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HKVHVL 8

Db 34 HKVHVL 39

; Patent No. 6210890
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN PEROXISOMAL THIOESTERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/265,294
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/100,851
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0293 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT09
; CLONE: 2150905
US-09-265-294-1

Query Match 68.0%; Score 34; DB 3; Length 311;
Best Local Similarity 71.4%; Pred. No. 94;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHVLV 9
| | | | |
Db 239 HKVHFV 245

RESULT 6
US-09-215-694-7
; Sequence 7, Application US/09215694B
; Patent No. 6391583
; GENERAL INFORMATION:
; APPLICANT: Wisconsin Alumni Research Foundation
; APPLICANT: Hutchinson, Charles R.
; APPLICANT: Kennedy, Jonathan n.m.i
; APPLICANT: Park, Cheonseok n.m.i
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
; FILE REFERENCE: 960296.95718
; CURRENT APPLICATION NUMBER: US/09/215,694B
; CURRENT FILING DATE: 1999-12-18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1068
; TYPE: PRT
; ORGANISM: Aspergillus terreus
US-09-215-694-7

Query Match 68.0%; Score 34; DB 4; Length 1068;
Best Local Similarity 71.4%; Pred. No. 3.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHY 7
| | | | |
Db 185 FLHRVHH 191

RESULT 7
US-09-328-352-7101
; Sequence 7101, Application US/09328352
; Patent No. 8562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7101
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7101

Query Match 66.0%; Score 33; DB 4; Length 144;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHVLV 9
| | | | |
Db 71 IHKAHYIL 78

RESULT 8
US-09-328-352-7331
; Sequence 7331, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7331
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7331

Query Match 66.0%; Score 33; DB 4; Length 157;
Best Local Similarity 44.4%; Pred. No. 73;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHVLV 9
| | | | |
Db 116 FMHEIHALI 124

RESULT 9
US-08-624-125-19
; Sequence 19, Application US/08624125
; Patent No. 5744341
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; APPLICANT: SUN, ZAIREN
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
; TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
; NUMBER OF SEQUENCES: 21

;; CORRESPONDENCE ADDRESS:
;; ADDRESSES: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;; CITY: ARLINGTON
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22202
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/624,125
;; FILING DATE: 29-MAR-1996
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KELBER, STEVEN B.
;; REGISTRATION NUMBER: 30,073
;; REFERENCE/DOCKET NUMBER: 2747-063-27
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-413-3000
;; TELEFAX: 703-413-2220
;; INFORMATION FOR SEQ ID NO: 19:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 288 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-624-125-19

Query Match 66.0%; Score 33; DB 1; Length 288;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHYL 8
||:|:|:
Db 189 FLNRIHYM 196

RESULT 10
US-08-937-155-19
; Sequence 19, Application US/08937155
; Patent No. 6524811
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; APPLICANT: SUN, ZAI-PEN
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
; TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,155
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/624,125
; FILING DATE: 29-MAR-1996
; ATTORNEY/AGENT INFORMATION:

;; NAME: KELBER, STEVEN B.
;; REGISTRATION NUMBER: 30,073
;; REFERENCE/DOCKET NUMBER: 2747-063-27
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-413-3000
;; TELEFAX: 703-413-2220
;; INFORMATION FOR SEQ ID NO: 19:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 288 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-937-155-19

Query Match 66.0%; Score 33; DB 4; Length 288;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHYL 8
||:|:|:
Db 189 FLNRIHYM 196

RESULT 11
US-09-126-420A-20
; Sequence 20, Application US/09126420A
; Patent No. 6376753
; GENERAL INFORMATION:
; APPLICANT: BATARD, YANNICK
; APPLICANT: ROBINEAU, TIBURCE
; APPLICANT: DURST, FRANCIS
; APPLICANT: WERCK-REICHAERT, DANIELE
; APPLICANT: DIDIERJEAN, LUC
; TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP76B1 FROM HELIANTHUS
; TITLE OF INVENTION: TUBEROSUS AND ITS APPLICATIONS AS BIOCATALYST IN
; TITLE OF INVENTION: PARTICULAR FOR THE DEGRADATION OF ENVIRONMENTAL
; TITLE OF INVENTION: POLLUTANTS AND FOR ALTERING THE RESISTANCE OF PLANTS S
; FILE REFERENCE: 03715.0032
; CURRENT APPLICATION NUMBER: US/09/126,420A
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: 60/054,351
; PRIOR FILING DATE: 1997-07-31
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 20
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Persea americana
; US-09-126-420A-20

Query Match 66.0%; Score 33; DB 4; Length 471;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYL 8
||:|:|:
Db 349 LHLHYL 355

RESULT 12
US-09-499-302A-7
; Sequence 7, Application US/09499302A
; Patent No. 6369212
; GENERAL INFORMATION:
; APPLICANT: BOUNG-JUN, OH
; APPLICANT: MOON, KYUNG KO
; APPLICANT: YOUNG, SOON KIM
; TITLE OF INVENTION: A CYTOCHROME P450 GENE HIGHLY EXPRESSED IN THE
; TITLE OF INVENTION: INCOMPATIBLE INTERACTION
; FILE REFERENCE: 10324/P64443USO
; CURRENT APPLICATION NUMBER: US/09/499,302A
; CURRENT FILING DATE: 2000-02-07

; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7

; LENGTH: 502
; TYPE: PRT

; ORGANISM: Persea sp.
; FEATURE:

; NAME/KEY: MOD_RES
; LOCATION: (1)

; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES

; LOCATION: (22)
; OTHER INFORMATION: variable or unknown amino acid

; NAME/KEY: MOD_RES
; LOCATION: (30)

; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES

; LOCATION: (49)
; OTHER INFORMATION: variable or unknown amino acid

; NAME/KEY: MOD_RES
; LOCATION: (199)

; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES

; LOCATION: (226)
; OTHER INFORMATION: variable or unknown amino acid

; NAME/KEY: MOD_RES
; LOCATION: (240)

; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES

; LOCATION: (247)
; OTHER INFORMATION: variable or unknown amino acid

; NAME/KEY: MOD_RES
; LOCATION: (262)

; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES

; LOCATION: (278)
; OTHER INFORMATION: variable or unknown amino acid

; US-09-499-302A-7

Query Match 66.0%; Score 33; DB 4; Length 502;
Best Local Similarity 71.4%; Pred. No. 2.2e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKVHYL 8

Db 349 LKQLHYL 355

RESULT 13

US-09-252-991A-23106

; Sequence 23106, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 23106

; LENGTH: 554

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23106

Query Match

Best Local Similarity 71.4%; Pred. No. 2.5e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKVHYL 8

Db 259 LKQLHYL 265

RESULT 14

US-08-190-802A-63

; Sequence 63, Application US/08190802A

; Patent No. 5513003

; GENERAL INFORMATION:

; APPLICANT: Mochly-Rosen, Daria

; APPLICANT: Ron, Dorit

; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses

; TITLE OF INVENTION: Thereof

; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dehlinger & Associates

; STREET: P.O. Box 60850

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94306-0850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/190,802A

; FILING DATE: 01-FEB-1994

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Fabian, Gary R.

; REGISTRATION NUMBER: 33,875

; REFERENCE/DOCKET NUMBER: 8600-0139

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 324-0880

; TELEFAX: (415) 324-0960

; INFORMATION FOR SEQ ID NO: 63:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 713 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; INDIVIDUAL ISOLATE: TUP1, Fig. 46

; US-08-190-802A-63

Query Match 66.0%; Score 33; DB 1; Length 713;

Best Local Similarity 75.0%; Pred. No. 3.1e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKVHYL 9

Db 279 LKQDHYL 286

RESULT 15

US-08-477-346-63

; Sequence 63, Application US/08477346

; Patent No. 6262023

; GENERAL INFORMATION:

; APPLICANT: Mochly-Rosen, Daria

; APPLICANT: Ron, Dorit

; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses

; TITLE OF INVENTION: Thereof

; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Morrison & Foerster

; STREET: 2000 Pennsylvania Avenue, NW

CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 713 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TUP1, Fig. 46
US-08-477-346-63

Query Match 66.0%; Score 33; DB 3; Length 713;
Best Local Similarity 75.0%; Pred. No. 3.le+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHYLV 9
||: ||||
Db 279 LQDHYLV 286

Search completed: December 16, 2003, 14:20:41
Job time : 13.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:10:15 ; Search time 10.3333 Seconds
(without alignments)
83.760 Million cell updates/sec

Title: US-09-870-089B-11

Perfect score: 50

Sequence: 1 FLHKVHYLV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB ID	Description
1	40	80.0	416	2 T47173	hypothetical prote
2	36	72.0	102	2 I30534	Ig kappa chain V r
3	36	72.0	113	2 H30534	Ig kappa chain V r
4	36	72.0	113	2 F30534	Ig kappa chain V r
5	36	72.0	114	1 KMS7A	Ig kappa chain V r
6	36	72.0	378	2 H85431	hypothetical prote
7	36	72.0	454	2 AF1217	Salmonella enteric
8	36	72.0	454	2 AI1570	Salmonella enteric
9	36	72.0	670	2 C71630	cytochrome C-type
10	36	72.0	1532	2 T18438	hypothetical prote
11	35	70.0	300	2 T38986	probable c-4 methy
12	35	70.0	347	2 T29415	hypothetical prote
13	35	70.0	650	2 T00094	endostyle-specific
14	35	70.0	662	2 H97834	cytochrome c-type
15	35	70.0	1645	2 T31339	carbamoyl-phosphat
16	34	68.0	58	2 C97788	hypothetical prote
17	34	68.0	122	2 H82332	ribosomal protein
18	34	68.0	125	2 G64076	ribosomal protein
19	34	68.0	130	2 AD0430	30S ribosomal prot
20	34	68.0	131	1 R38C6	ribosomal protein
21	34	68.0	131	2 H86116	30S ribosomal subu
22	34	68.0	131	2 H91275	30S ribosomal subu
23	34	68.0	131	2 AG1052	30S ribosomal subu
24	34	68.0	200	1 H64062	membrane-bound tet
25	34	68.0	261	2 G72470	probable ABC trans
26	34	68.0	303	2 AE1066	hypothetical prote
27	34	68.0	319	1 JC5644	acyl-CoA thioleste
28	34	68.0	365	2 JQ1146	C-5 sterol desatur
29	34	68.0	514	2 T10559	hypothetical prote

30	34	68.0	525	2 T10574	hypothetical prote
31	34	68.0	545	2 T02578	hypothetical prote
32	34	68.0	569	2 T03390	4-coumarate-CoA li
33	34	68.0	651	2 E85024	probable CHP-rich
34	34	68.0	652	2 G85024	probable CHP-rich
35	34	68.0	652	2 T02001	hypothetical prote
36	34	68.0	658	2 F85024	probable CHP-rich
37	34	68.0	668	2 T10575	hypothetical prote
38	34	68.0	989	2 T01519	hypothetical prote
39	33	66.0	53	2 C82714	50S ribosomal prot
40	33	66.0	288	2 A34440	isopentenyl-diphos
41	33	66.0	345	2 T29416	hypothetical prote
42	33	66.0	471	1 A35867	cytochrome P450 71
43	33	66.0	473	2 T31717	hypothetical prote
44	33	66.0	474	2 T34193	G protein-coupled
45	33	66.0	502	2 T52256	cytochrome P-450LX

ALIGNMENTS

RESULT 1

T47173

hypothetical protein DKFp7621156.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47173

R:Blöcker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24376

A:Accession: T47173

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-416 <AAA>

A:Cross-references: EMBL:AL162071

A:Experimental source: adult melanoma (MeWo cell line); clone DKFp7621156

C:Genetics:

A>Note: DKFp7621156.1

Query Match 80.0%; Score 40; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYL 8
DB 275 LHKVHYL 281

RESULT 2

I30534

Ig kappa chain V region (224.4C9) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 10-Feb-1989 #sequence_revision 10-Feb-1989 #text_change 21-Jan-2000

C:Accession: I30534

R:Clafin, J.L.; Berry, J.

J. Immunol. 141, 4012-4019, 1988

A:Title: Genetics of the phosphocholine-specific antibody response to Streptococcus pne

A:Reference number: A30534; MUID:89035545; PMID:3141511

A:Accession: I30534

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-102 <CLA>

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:16-96/Domain: immunoglobulin homology <IMM>

Query Match 72.0%; Score 36; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHYL 8
DB 34 HKVHYL 39

RESULT 3

H30534
IG kappa chain V region (224.4B11) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 10-Feb-1989 #sequence_revision 10-Feb-1989 #text_change 21-Jan-2000
C;Accession: H30534
R;Claflin, J.L.; Berry, J.
J. Immunol. 141, 4012-4019, 1988
A;Title: Genetics of the phosphocholine-specific antibody response to Streptococcus pneumoniae
A;Reference number: A30534; MUID:89035545; PMID:3141511
A;Accession: H30534
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-113 <CLA>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotrimer; immunoglobulin
F;16-96/Domain: immunoglobulin homology <IMM>

Query Match 72.0%; Score 36; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHYL 8

|||||

DB 34 HKVHYL 39

RESULT 4

F30534
IG kappa chain V region (224.2B11) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 10-Feb-1989 #sequence_revision 10-Feb-1989 #text_change 21-Jan-2000
C;Accession: F30534
R;Claflin, J.L.; Berry, J.
J. Immunol. 141, 4012-4019, 1988
A;Title: Genetics of the phosphocholine-specific antibody response to Streptococcus pneumoniae
A;Reference number: A30534; MUID:89035545; PMID:3141511
A;Accession: F30534
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-113 <CLA>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotrimer; immunoglobulin
F;16-96/Domain: immunoglobulin homology <IMM>

Query Match 72.0%; Score 36; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHYL 8

|||||

DB 34 HKVHYL 39

RESULT 5

KVMS7A
IG kappa chain V region (S107A) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Feb-1984 #sequence_revision 19-Feb-1984 #text_change 13-Mar-1997
C;Accession: A01915; B30538; D30534; E30534; A30538; C30538; D30538
R;Kwan, S.P.; Rudikoff, S.; Seidman, J.G.; Leder, P.; Scharff, M.D.
J. Exp. Med. 153, 1366-1370, 1981
A;Title: Nucleic acid and protein sequences of phosphocholine-binding light chains.
A;Reference number: A01915; MUID:81241357; PMID:6788890
A;Accession: A01915
A;Molecule type: mRNA
A;Residues: 1-114 <KWA>
A;Note: the sequence was determined from the differentiated gene
R;Claflin, J.L.; Berry, J.
J. Immunol. 141, 4012-4019, 1988
A;Title: Genetics of the phosphocholine-specific antibody response to Streptococcus pneu

A;Reference number: A30534; MUID:89035545; PMID:3141511
A;Accession: B30538
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-106 <CLA>
A;Experimental source: clone 252.5E10
A;Accession: D30534
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-104 <CL2>
A;Experimental source: clone 224.2B8
A;Accession: E30534
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-113 <CL3>
A;Experimental source: clone 224.2G3
A;Accession: G30534
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-105 <CL4>
A;Experimental source: clone 224.1C7
A;Accession: A30538
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 17-113 <CL5>
A;Experimental source: clone 224.7E7
A;Accession: C30538
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 26-112 <CL6>
A;Experimental source: clone 252.5E11
A;Accession: D30538
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 32-104 <CL7>
A;Experimental source: 253.12G5
C;Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kap
hain disulfide bonds. In some cases, such as IGA and IGM, the subunits associate into la
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotrimer; immunoglobulin
F;16-96/Domain: immunoglobulin homology <IMM>
F;23-94/Disulfide bonds: #status predicted

Query Match 72.0%; Score 36; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHYL 8

|||||

DB 34 HKVHYL 39

RESULT 6

C85431
hypothetical protein AT4936530 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: C85431
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprir
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: C85431
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-378 <STO>
A;Cross-references: GB:NC_001268; NID:g7270601; PIDN:CAB80319.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4936530
A;Map position: 4

Query Match 72.0%; Score 36; DB 2; Length 378;
Best Local Similarity 71.4%; Pred. No. 25;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HKVHVILV 9
||:||||
Db 89 HKIHVVV 95

RESULT 7
AF1217
Salmonella enterica pdsu protein homolog lm0142 [imported] - Listeria monocytogenes (strain
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AF1217
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
ok, C.; Schluster, T.; Simeos, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1217
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-454 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99220.1; PID:g16410558; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lm0142

Query Match 72.0%; Score 36; DB 2; Length 454;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLHKVHY 7
:|||||
Db 239 YLKLHY 245

RESULT 8
A11570
Salmonella enterica pdsu protein homolog lin1106 [imported] - Listeria innocua (strain C
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: A11570
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
ok, C.; Schluster, T.; Simeos, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: A11570
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-454 <GLA>
A;Cross-references: GB:AL592022; PIDN:CA96337.1; PID:g16413565; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin1106

Query Match 72.0%; Score 36; DB 2; Length 454;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLHKVHY 7
:|||||
Db 239 YLKLHY 245

RESULT 9
C71630

cytochrome C-type biogenesis protein ccmF (ccmf) RP703 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C;Accession: C71630
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichertitz-Ponten, T.; Alsmark, I.
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499; PMID:9823893
A;Accession: C71630
A;Molecule type: DNA
A;Residues: 1-670 <AND>
A;Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAAL51139.1; PID:g38612.
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: ccmF; RP703
C;Superfamily: nrfE protein

Query Match 72.0%; Score 36; DB 2; Length 670;
Best Local Similarity 75.0%; Pred. No. 47;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKVHVILV 9
||:||||
Db 500 LHNKVHYLL 507

RESULT 10
T18438
hypothetical protein C0415c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2002
C;Accession: T18438
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18438
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1532 <LAW>
A;Cross-references: EMBL:Z98547; NID:g1325376; PIDN:CAB11117.1
C;Genetics:
A;Map position: 3
A;Note: C0415c

Query Match 72.0%; Score 36; DB 2; Length 1532;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLHKVHVILV 9
||:||||
Db 696 FLNKVHYSI 704

RESULT 11
T38986
Probable C-4 methyl sterol oxidase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T38986
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Zimmermann, W.; Wambutt, R.
submitted to the EMBL Data Library, August 1999
A;Reference number: Z21822
A;Accession: T38986
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-300 <MCD>
A;Cross-references: EMBL:AL109832; PIDN:CAB52730.1; GSPDB:GN00066; SPDB:SPAC630.08c
A;Experimental source: strain 972h-; cosmid c630
C;Genetics:
A;Gene: SPDB:SPAC630.08c
A;Map position: 1
C;Superfamily: Saccharomyces cerevisiae ERG25 protein

Query Match 70.0%; Score 35; DB 2; Length 300;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHY 7
|:|||||:
Db 165 FLHKVHH 171

RESULT 12
T29415
hypothetical protein F21F8.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29415
R;Wilson, R.; Favello, A.; Le, T.T.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid F21F8.
A:Reference number: Z20618
A:Accession: T29415
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-347 <WIL>
A:Cross-references: EMBL:U97000; PIDN:AAC47999.1; GSPDB:GN00023; CESP:F21F8.9
A:Experimental source: strain Bristol N2; clone F21F8
C:Genetics:
A:Gene: CESP:F21F8.9
A:Map position: 5
A:Introns: 67/3; 97/1; 123/2; 215/3; 258/3; 318/1

Query Match 70.0%; Score 35; DB 2; Length 347;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHYL 8
|:|||||:
Db 10 YLHSIHYI 17

RESULT 13
T00094
endostyle-specific protein - sea squirt (Ciona intestinalis)
C:Species: Ciona intestinalis
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 23-Mar-2001
C:Accession: T00094
R;Ogasawara, M.; Satoh, N.
submitted to the EMBL Data Library, February 1998
A:Description: Isolation and characterization of endostyle-specific genes in the Ascidia
A:Reference number: Z14109
A:Accession: T00094
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-650 <OGA>
A:Cross-references: EMBL:AB010895
A:Experimental source: endostyle

Query Match 70.0%; Score 35; DB 2; Length 650;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVH 6
|:|||||:
Db 259 FLHKVH 264

RESULT 14
H97834
cytochrome c-type biogenesis protein ccmf [imported] - Rickettsia conorii (strain Malish)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: H97834
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rd

Science 293, 2093-2098, 2001
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A;Reference number: A97700; MUID:21442074; PMID:11557893
A;Accession: H97834
A;Status: preliminary
A:Molecule type: DNA
A:Residues: 1-662 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03618.1; PID:g15620202; GSPDB:GN00173
C:Genetics:
A:Gene: ccmf
C;Superfamily: nrfE protein

Query Match 70.0%; Score 35; DB 2; Length 662;
Best Local Similarity 62.5%; Pred. No. 72;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHYLV 9
|:|||||:
Db 492 LNHVYLL 499

RESULT 15
T31339
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) III - Babesia bovis
C:Species: Babesia bovis
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
C:Accession: T31339
R;Bagnara, A.S.; Chansiri, K.
Gene 172, 173-174, 1996
A;Title: Sequences upstream and downstream from the glutamine-dependent carbamoyl phosph
A:Reference number: Z21005; MUID:96257278; PMID:8654985
A:Accession: T31339
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1645 <BAG>
A:Cross-references: EMBL:U18792; MID:g944893; PID:g619743; PIDN:AAC47302.1
C:Genetics:
A:Gene: CPSII
C;Superfamily: carbamoyl-phosphate synthase (ammonia); biotin carboxylase homology; carb
omology; carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology; trpG
C;Keywords: ligase
F;26-1623/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>

Query Match 70.0%; Score 35; DB 2; Length 1645;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHYL 8
|:|||||:
Db 970 FLHRLHLL 977

Search completed: December 16, 2003, 14:19:18
Job time : 11.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:35 ; Search time 5.66667 Seconds
(without alignments)
74.689 Million cell updates/sec

Title: US-09-870-089b-11
Perfect score: 50
Sequence: 1 FLKHVHLYV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	80.0	735	1 MKLN HUMAN	Q9ul63 homo sapien
2	40	80.0	735	1 MKLN MOUSE	Q99050 mus musculus
3	40	80.0	735	1 MKLN RAT	Q990v3 rattus norv
4	36	72.0	114	1 KVIA_MOUSE	P01632 mus musculus
5	35	70.0	113	1 RS6_EUCAP	Q8x918 buchnera ap
6	35	70.0	300	1 ER35_SCHPO	Q9uuu4 schizosacch
7	34	68.0	122	1 RS6_VIBCH	Q9kuz2 vibrio chol
8	34	68.0	123	1 RS6_VIBVU	Q8dc16 vibrio ynf
9	34	68.0	125	1 RS6_HASIN	P44375 haemophilus
10	34	68.0	125	1 RS6_PASMU	Q9c1n8 pascalella
11	34	68.0	129	1 RS6_VIBPA	Q87172 vibrio para
12	34	68.0	130	1 RS6_YERPE	Q8zbg2 yersinia pe
13	34	68.0	131	1 RS6_ECOL6	Q8xd11 escherichia
14	34	68.0	131	1 RS6_ECOL6	Q8xg88 salmonella
15	34	68.0	135	1 RS6_ECOL1	P02358 escherichia
16	34	68.0	200	1 NACP_HAEIN	P44655 haemophilus
17	34	68.0	219	1 PTEL_HUMAN	O14734 homo sapien
18	34	68.0	365	1 PTEL_YEAST	P32353 saccharomyc
19	34	68.0	569	1 ACL2_ORYSA	Q42982 oryza sativ
20	34	68.0	597	1 GAN_HUMAN	Q9h2c0 homo sapien
21	34	68.0	697	1 PURL_PYRAE	Q8zzj7 pyrobaculum
22	34	68.0	1043	1 RRPO_NODAV	Q9imn4 nodamura vi
23	34	68.0	1068	1 HMDB_ASPTT	Q9y7d2 aspergillus
24	33	66.0	119	1 RS6_EUCBP	Q89441 buchnera ap
25	33	66.0	287	1 IDI1_YEAST	P15496 saccharomyc
26	33	66.0	316	1 RPOA_MESVI	Q9muv0 mesostigma
27	33	66.0	364	1 ERG3_CANGA	P50860 candida gla
28	33	66.0	447	1 DHE4_PSYT1	Q981f9 psychrobact
29	33	66.0	454	1 NFM_FIG	P08552 sus scrofa
30	33	66.0	471	1 CP71_PRAE	P24465 persea amer
31	33	66.0	515	1 G6PD_TREPA	O83491 treponema p
32	33	66.0	544	1 C821_PEA	Q43068 pisum sativ
33	33	66.0	713	1 TUPI_YEAST	P16649 saccharomyc

ALIGNMENTS

RESULT 1				
MKLN HUMAN				
ID	MKLN HUMAN	STANDARD;	PRT;	735 AA.
AC	Q9UL63; Q9NSK4; Q9NUS8;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Muskelin.			
GN	MKLN1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain, Kidney, Lung, and Osteosarcoma;			
RX	MEDLINE=20108780; PubMed=10640805;			
RA	Adams J.C., Zhang L.;			
RT	"cDNA cloning of human muskelin and localisation of the muskelin			
RT	(MKLN1) gene to human chromosome 7q32 and mouse chromosome 6 B1/B2 by			
RT	Physical mapping and FISH."			
RL	Cytogenet. Cell Genet. 87:19-21(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,			
RA	Tanai H., Kimata M., Watanabe M., Hiraoaka S., Ishii S., Kawai Y.,			
RA	Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,			
RA	Masuko Y., Kanehori K.;			
RT	"NEO human cDNA sequencing project."			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 186-735 FROM N.A.			
RC	TISSUE=Uterus;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins P.S., Wagner K.H., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			
RA	Bosak S.S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón A.C., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,			
RA	Schurich A., Schein J.E., Jones S.J.M., Marra M.A.,			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			

P12839 rattus norv
P08553 mus musculus
P16053 gallus gall
O78441 guillardi
P07197 homo sapien
P73505 synchocyst
P41959 brugia paha
O07813 neisseria g
Q9J024 neisseria m
Q9J429 neisseria m
Q10132 schizosacch
P19654 bacterioph


```

RN RP SEQUENCE OF 320-735 FROM N.A.
RC TISSUE=Melanoma;
RA Bloeker H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Acts as a mediator of cell spreading and cytoskeletal
CC responses to the extracellular matrix component thrombospondin 1
CC (By similarity).
CC -!- SUBUNIT: Binds with the C-terminal tail of the prostaglandin EP3
CC receptor alpha isoform (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Contains 5 Kelch repeats.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 1 CTLD domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC stop codon in position 171.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF047489; AAF06698.1; -.
DR EMBL; AK002024; BAA92042.1; ALT SEQ.
DR EMBL; BC002834; AAH02834.1; ALT_INIT.
DR EMBL; AL162071; CAB82407.1; -.
DR PIR; T47173; T47173.
DR MIM; 605623; -.
DR Genew; HGNC:7109; MKLN1.
DR InterPro; IPR006595; CTLD.
DR InterPro; IPR006552; Kelch_rep.
DR InterPro; IPR006594; Lish.
DR Pfam; PF01344; Kelch; 3.
DR SMART; SM00668; CTLD; 1.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS50897; CTLD; 1.
DR PROSITE; PS50896; Lish; 1.
KW Kelch repeat; Repeat.
FT DOMAIN 172 204 LISH.
FT REPEAT 206 258 CTLD.
FT REPEAT 284 330 KELCH 1.
FT REPEAT 339 391 KELCH 2.
FT REPEAT 408 458 KELCH 3.
FT REPEAT 469 515 KELCH 4.
FT REPEAT 526 578 KELCH 5.
FT CONFLICT 6 6 A -> M (IN REF. 1).
FT CONFLICT 111 111 N -> Y (IN REF. 1).
FT CONFLICT 408 408 I -> N (IN REF. 1).
FT CONFLICT 414 414 S -> G (IN REF. 1).
FT CONFLICT 436 436 Q -> K (IN REF. 1).
FT CONFLICT 474 474 G -> A (IN REF. 1).
FT CONFLICT 502 502 T -> N (IN REF. 1).
SQ SEQUENCE 735 AA; 84767 MW; 1A8F06FE1DE9193D CRC64;

Query Match 80.0%; Score 40; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYL 8
Db 594 LHKVHYL 600

RESULT 2
MKLN_MOUSE STANDARD; PRT; 735 AA.
AC O89050;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

```

```

DE Muskelin.
GN MKLN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98393538; PubMed=9724633;
RA Adams J.C., Seed B., Lawler J.;
RT "Muskelin, a novel intracellular mediator of cell adhesive and
RT cytoskeletal responses to thrombospondin-1.";
RL EMBO J. 17:4964-4974 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Breast tumor;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toehiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Acts as a mediator of cell spreading and cytoskeletal
CC responses to the extracellular matrix component thrombospondin 1.
CC -!- SUBUNIT: Binds with the C-terminal tail of the prostaglandin EP3
CC receptor alpha isoform (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Contains 5 Kelch repeats.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 1 CTLD domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U72194; AAC63849.1; -.
DR EMBL; BC013703; AAH13703.1; -.
DR MGD; MGI:1351638; Mklm1.
DR InterPro; IPR006595; CTLD.
DR InterPro; IPR006652; Kelch_rep.
DR InterPro; IPR006594; Lish.
DR Pfam; PF01344; Kelch; 3.
DR SMART; SM00668; CTLD; 1.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS50897; CTLD; 1.
DR PROSITE; PS50896; Lish; 1.
KW Kelch repeat; Repeat.
FT DOMAIN 172 204 LISH.
FT DOMAIN 206 258 CTLD.
FT REPEAT 284 330 KELCH 1.
FT REPEAT 339 391 KELCH 2.
FT REPEAT 400 458 KELCH 3.
FT REPEAT 469 515 KELCH 4.
FT REPEAT 526 578 KELCH 5.

```

```
SQ SEQUENCE 735 AA; 84877 MW; 7B0C8D2CB8C242AA CRC64;
Query Match 80.0%; Score 40; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYL 8
|||||
Db 594 LHKVHYL 600

RESULT 3
MKLN RAT STANDARD; PRT; 735 AA.
AC Q99PV3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE MusKelin.
GN MKLIN OR MSK.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND INTERACTION WITH EP3-ALPHA.
RX MEDLINE=20462961; PubMed=11006128;
RA Hasegawa H., Katoh H., Fujita H., Mori K., Negishi M.;
RT "Receptor isoform-specific interaction of prostaglandin EP3 receptor with muskelin.";
RT with muskelin.";
RL Biochem. Biophys. Res. Commun. 276:350-354(2000).
CC -!- FUNCTION: Acts as a mediator of cell spreading and cytoskeletal responses to the extracellular matrix component thrombospondin 1 (By similarity).
CC -!- SUBUNIT: Binds with the C-terminal tail of the prostaglandin EP3 receptor alpha isoform.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Contains 5 Kelch repeats.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 1 CTLD domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB046442; BAB21439.1; -.
DR InterPro; IPR006595; CTLD.
DR InterPro; IPR006652; Kelch_rep.
DR Pfam; PF01344; Kelch; 3.
DR SMART; SM00668; CTLD; 1.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS50897; CTLD; 1.
DR PROSITE; PS50896; Lish; 1.
KW Kelch repeat; Repeat.
FT DOMAIN 172 204 LISH.
FT DOMAIN 206 258 CTLD.
FT REPEAT 284 330 KELCH 1.
FT REPEAT 339 391 KELCH 2.
FT REPEAT 400 458 KELCH 3.
FT REPEAT 469 515 KELCH 4.
FT REPEAT 526 578 KELCH 5.
SQ SEQUENCE 735 AA; 84833 MW; 79BD0B8F74419E76 CRC64;

Query Match 80.0%; Score 40; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYL 8

SQ SEQUENCE 735 AA; 84877 MW; 7B0C8D2CB8C242AA CRC64;
Query Match 80.0%; Score 40; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYL 8
|||||
Db 594 LHKVHYL 600

RESULT 4
KVIA MOUSE STANDARD; PRT; 114 AA.
AC P01632;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-I region SI07A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81241357; PubMed=6788890;
RA Kwan S.-P., Rudikoff S., Seidman J.G., Leder P., Scharff M.D.;
RT "Nucleic acid and protein sequences of phosphocholine-binding light chains.";
RL J. Exp. Med. 153:1366-1370(1981).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29423; AAC00033.1; -.
DR PIR; A01915; KVMS7A.
DR HSSP; P01607; IREI.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 40 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 41 55 FRAMEWORK-2.
FT DOMAIN 56 62 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 63 94 FRAMEWORK-3.
FT DOMAIN 95 103 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 104 113 FRAMEWORK-4.
FT DISULFID 123 94 BY SIMILARITY.
FT NON_TER 114 114
SQ SEQUENCE 114 AA; 12717 MW; 32008EC8E9DBE67B CRC64;

Query Match 72.0%; Score 36; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHYL 8
|||||
Db 34 HKVHYL 39

RESULT 5
RS6 BUCAP STANDARD; PRT; 113 AA.
AC Q8K318;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S6.
GN RPSF OR BUSG544.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
```

```

OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamás I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Werngreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002)
CC -!- FUNCTION: Binds together with S18 to 16S ribosomal RNA (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE SGP FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB014128; AAM68083.1; --
CC HAMAP; MF_00360; --; 1.
CC InterPro; IPR000529; Ribosomal_S6.
CC Pfam; PF01250; Ribosomal_S6; 1.
CC ProDom; PD003809; Ribosomal_S6; 1.
CC TIGRfam; TIGR00166; S6; 1.
CC PROSITE; PS01048; RIBOSOMAL_S6; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
CC S6 SEQUENCE 113 AA; 13501 MW; DEID2B6344CIBEC CRC64;
DR -----
DR Query Match 70.0%; Score 35; DB 1; Length 113;
DR Best Local Similarity 62.5%; Pred. No. 5.1;
DR Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LHKVHYLV 9
Db 54 LHKAHYIL 61
||| |||:
||| |||:

RESULT 6
ER25 SCHPO STANDARD; PRT; 300 AA.
ID ER25 SCHPO STANDARD; PRT; 300 AA.
AC Q9U0H4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE C-4 methyl sterol oxidase [EC 1.-.-.-].
GN ERG25 OR SPAC630.08C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds R., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Hoizer E., Moestl D., Hilbert H.,

```

```

RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Ruvetta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Spakovski G.V., Ussery D., Bartell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Catalyzes the first step in the removal of the two C-4
CC methyl groups of 4,4-dimethylzymosterol (By similarity).
CC -!- COFACTOR: Iron (By similarity).
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum. Membrane-associated
CC (Probable).
CC -!- DOMAIN: The histidine box domains may contain the active site
CC and/or be involved in metal ion binding.
CC -!- SIMILARITY: BELONGS TO THE STEROL DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL109832; CA52730.1; --
CC PIR; T38986; T38986.
CC GeneDB SPombe; SPAC630.08C; --
CC InterPro; IPR006087; Sterol_desat.
CC InterPro; IPR006088; Sterol_desatur.
CC Pfam; PF01598; Sterol_desat; 1.
CC Sterol biosynthesis; Oxidoreductase; Endoplasmic reticulum;
CC Transmembrane; Iron.
CC TRANSMEM 186 206 POTENTIAL.
CC FT DOMAIN 154 158 HISTIDINE BOX-1.
CC FT DOMAIN 167 171 HISTIDINE BOX-2.
CC FT DOMAIN 251 257 HISTIDINE BOX-3.
CC SEQUENCE 300 AA; 35802 MW; AA8C2603D35DF331 CRC64;
DR -----
DR Query Match 70.0%; Score 35; DB 1; Length 300;
DR Best Local Similarity 71.4%; Pred. No. 14;
DR Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 FLKHVHY 7
Db 165 FIHKVHH 171
||: |||:
||: |||:

RESULT 7
RS6_VIBCH STANDARD; PRT; 122 AA.
ID RS6_VIBCH STANDARD; PRT; 122 AA.
AC Q9XU22;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S6.
GN RPS6 OR VC0366.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,

```

McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;
"DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.";
RL Nature 406:477-483(2000)
CC -!- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE004124; AAF93539.1; -.
CC PIR; H82332; H82332.
CC TIGR; V00366; -.
CC HAMAP; MF_00360; -; 1.
CC InterPro; IPR000529; Ribosomal_S6.
CC Pfam; PF01250; Ribosomal_S6; 1.
CC ProDom; PD003809; Ribosomal_S6; 1.
CC TIGRFAMs; TIGR00166; S6; 1.
CC PROSITE; PS01048; RIBOSOMAL_S6; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
KW RIBOSOMAL protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 122 AA; 14249 MW; 4BA3DB6B1FDDBA01 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 122;
Best Local Similarity 62.5%; Pred. No. 8.6;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

CC -!- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016801; AAC09836.1; -.
CC HAMAP; MF_00360; -; 1.
CC InterPro; IPR000529; Ribosomal_S6.
CC Pfam; PF01250; Ribosomal_S6; 1.
CC ProDom; PD003809; Ribosomal_S6; 1.
CC TIGRFAMs; TIGR00166; S6; 1.
CC PROSITE; PS01048; RIBOSOMAL_S6; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
KW RIBOSOMAL protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 125 AA; 14484 MW; A3804A2FEPBE8EAC CRC64;

Query Match 68.0%; Score 34; DB 1; Length 125;
Best Local Similarity 62.5%; Pred. No. 8.6;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

CC -!- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32736; AAC22205.1; -.
CC PIR; G64076; G64076.
CC TIGR; H10547; -.
CC HAMAP; MF_00360; -; 1.
CC InterPro; IPR000529; Ribosomal_S6.
CC Pfam; PF01250; Ribosomal_S6; 1.
CC ProDom; PD003809; Ribosomal_S6; 1.
CC TIGRFAMs; TIGR00166; S6; 1.
CC PROSITE; PS01048; RIBOSOMAL_S6; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
KW RIBOSOMAL protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 125 AA; 14484 MW; A3804A2FEPBE8EAC CRC64;

DR InterPro; IPR000529; Ribosomal_S6.
DR Pfam; PF01250; Ribosomal_S6; 1.
DR ProDom; PD003809; Ribosomal_S6; 1.
DR TIGRFAMs; TIGR00166; S6; 1.
DR PROSITE; PS01048; RIBOSOMAL_S6; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 123 AA; 14301 MW; FCA3C5582FDA3C03 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 123;
Best Local Similarity 62.5%; Pred. No. 8.6;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

CC -!- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32736; AAC22205.1; -.
CC PIR; G64076; G64076.
CC TIGR; H10547; -.
CC HAMAP; MF_00360; -; 1.
CC InterPro; IPR000529; Ribosomal_S6.
CC Pfam; PF01250; Ribosomal_S6; 1.
CC ProDom; PD003809; Ribosomal_S6; 1.
CC TIGRFAMs; TIGR00166; S6; 1.
CC PROSITE; PS01048; RIBOSOMAL_S6; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
KW RIBOSOMAL protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 125 AA; 14484 MW; A3804A2FEPBE8EAC CRC64;

Query Match 68.0%; Score 34; DB 1; Length 125;
Best Local Similarity 62.5%; Pred. No. 8.6;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

CC -!- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32736; AAC22205.1; -.
CC PIR; G64076; G64076.
CC TIGR; H10547; -.
CC HAMAP; MF_00360; -; 1.
CC InterPro; IPR000529; Ribosomal_S6.
CC Pfam; PF01250; Ribosomal_S6; 1.
CC ProDom; PD003809; Ribosomal_S6; 1.
CC TIGRFAMs; TIGR00166; S6; 1.
CC PROSITE; PS01048; RIBOSOMAL_S6; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
KW RIBOSOMAL protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 125 AA; 14484 MW; A3804A2FEPBE8EAC CRC64;

```

Best Local Similarity 62.5%; Pred. No. 8.8;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHVL 9
DB 54 LKHAHYVL 61

RESULT 10
RS6_PASMU STANDARD; PRT; 125 AA.
AC Q9CLN8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S6.
GN RPSF OR RPS6 OR PM1180.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC [1]
DR EMBL; AB006158; AAK03264.1; -.
DR HAMAP; MF_00360; -.
DR InterPro; IPR000529; Ribosomal_S6.
DR Pfam; PF01250; Ribosomal_S6; 1.
DR ProDom; PD003809; Ribosomal_S6; 1.
DR TIGRFAMs; TIGR00166; S6; 1.
DR PROSITE; PS01048; RIBOSOMAL_S6; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 125 AA; 14505 MW; 102C842FD3FF7736 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 125;
Best Local Similarity 62.5%; Pred. No. 8.8;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHVL 9
DB 54 LKHAHYVL 61

RESULT 11
RS6_VIBPA STANDARD; PRT; 129 AA.
AC Q87L72;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S6.
GN RPSF OR VP2740.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Versinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Versinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nejima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -!- FUNCTION: Binds together with S18 to 16S ribosomal RNA (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC [1]
DR EMBL; AP005082; BAC61003.1; -.
DR HAMAP; MF_00360; -.
DR PROSITE; PS01048; RIBOSOMAL_S6; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 129 AA; 14983 MW; 63AB159EC2B9F561 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 129;
Best Local Similarity 62.5%; Pred. No. 9.1;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHVL 9
DB 54 LKHAHYVL 61

RESULT 12
RS6_YERPE STANDARD; PRT; 130 AA.
AC Q8ZB81;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S6.
GN RPSF OR YPO3539 OR Y0645.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaihia M., James K.D., Brooks K., Cerdano-Tarraga A.M.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Davis P., Dougan G.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Versinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).

```

```

CC  -!- FUNCTION: Binds together with S18 to 16S ribosomal RNA (By
CC  similarity).
CC  -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AJ414157; CAC92768.1; -
DR  EMBL; AE013667; AM84233.1; -
DR  FIR; AD0430; AD0430.
DR  HAMAP; MF_00360; -; 1.
DR  InterPro; IPR000529; Ribosomal_S6.
DR  Pfam; PF01250; Ribosomal_S6; 1.
DR  ProDom; PD003809; Ribosomal_S6; 1.
DR  TIGRPFAMs; TIGR00166; S6; 1.
DR  PROSITE; PS01048; RIBOSOMAL_S6; 1.
DR  Ribosomal protein; rRNA-binding; Complete proteome.
DR  NCBI_TaxID=15008 MW; 7897E436DD1FA562 CRC64;
SQ  SEQUENCE 130 AA; 15008 MW; 7897E436DD1FA562 CRC64;

Query Match      68.0%; Score 34; DB 1; Length 130;
Best Local Similarity 62.5%; Pred. No. 9.2;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY  2 LHKVHYLV 9
   ||| |||:
Db  54 LHKAHYVL 61

RESULT 13
RS6_ECOL6
ID  RS6_ECOL6 STANDARD; PRT; 131 AA.
AC  O8XDI1.
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  30S ribosomal protein S6.
GN  RPSF OR C5291 OR Z5809 OR ECS5176 OR SF4354.
OS  Escherichia coli O6.
OS  Shigella flexneri.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Escherichia.
OX  NCBI_TaxID=217992, 83334, 623;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX  MEDLINE=22388234; PubMed=12471157;
RA  Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA  Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA  Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA  Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT  "Extensive mosaic structure revealed by the complete genome sequence
RT  of uropathogenic Escherichia coli.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX  MEDLINE=21074935; PubMed=11206551;
RA  Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA  Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA  Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA  Welch R.A., Blattner F.R.;
RT  "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL  Nature 409:529-533 (2001).
RN  [3]
RP  SEQUENCE FROM N.A.

```

```

RC  SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
RX  MEDLINE=21156231; PubMed=11258796;
RA  Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22 (2001).
RN  [4]
RP  SEQUENCE FROM N.A.
RC  SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX  MEDLINE=22272406; PubMed=12384590;
RA  Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA  Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA  Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA  Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA  Yu J.;
RT  "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT  through comparison with genomes of Escherichia coli K12 and O157.";
RL  Nucleic Acids Res. 30:4432-4441 (2002).
CC  -!- FUNCTION: Binds together with S18 to 16S ribosomal RNA (By
CC  similarity).
CC  -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE016771; AAN83712.1; -
DR  EMBL; AE005652; AAG59396.1; -
DR  EMBL; AP002568; BAB38599.1; -
DR  EMBL; AE015442; AAN45771.1; -
DR  FIR; C65231; R3EC6.
DR  FIR; H86116; H86116.
DR  FIR; H91275; H91275.
DR  HAMAP; MF_00360; -; 1.
DR  InterPro; IPR000529; Ribosomal_S6.
DR  Pfam; PF01250; Ribosomal_S6; 1.
DR  ProDom; PD003809; Ribosomal_S6; 1.
DR  TIGRPFAMs; TIGR00166; S6; 1.
DR  PROSITE; PS01048; RIBOSOMAL_S6; 1.
DR  Ribosomal protein; rRNA-binding; Complete proteome.
DR  NCBI_TaxID=131 AA; 15187 MW; 01C1FD0EP4197ED2 CRC64;
SQ  SEQUENCE 131 AA; 15187 MW; 01C1FD0EP4197ED2 CRC64;

Query Match      68.0%; Score 34; DB 1; Length 131;
Best Local Similarity 62.5%; Pred. No. 9.2;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY  2 LHKVHYLV 9
   ||| |||:
Db  54 LHKAHYVL 61

RESULT 14
RS6_SALTY
ID  RS6_SALTY STANDARD; PRT; 131 AA.
AC  O8XG86;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  30S ribosomal protein S6.
GN  RPSF OR STM4391 OR STY4747 OR T4442.
OS  Salmonella typhimurium, and
OS  Salmonella typhi.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Salmonella.
OX  NCBI_TaxID=602, 601,
RN  [1]

```

```

RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher J., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337 (2003).
CC -1- FUNCTION: Binds together with S18 to 16S ribosomal RNA (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE SEP FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE008906; AAL23211.1; -
CC EMBL; AL627283; CAD06868.1; -
CC EMBL; AE016849; AA071899.1; -
CC StyGene; SG????; tpsf.
CC HAMAP; MF_00360; -; 1.
CC InterPro; IPR000529; Ribosomal_S6.
CC Pfam; PF01250; Ribosomal_S6; 1.
CC ProDom; PD003809; Ribosomal_S6; 1.
CC TIGRFAMS; TIGR00166; S6; 1.
CC PROSITE; PS01048; RIBOSOMAL_S6; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 131 AA; 15173 MW; 54974D6342C21362 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 131;
Best Local Similarity 62.5%; Pred. No. 9.2;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKVHYLV 9
||| |||
Db 54 LHKVHYLV 61

RESULT 15
RS6_ECOLI
ID -RS6_ECOLI STANDARD; PRT; 135 AA.
AC P02358;

```

```

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S6.
GN RPSF OR B4200.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE OF 1-131 FROM N.A.
RX MEDLINE=86310297; PubMed=3528756;
RA Schmier J., Kitakawa M., Isono K.;
RT "The nucleotide sequence of an Escherichia coli chromosomal region
RT containing the genes for ribosomal proteins S6, S18, L9 and an open
RT reading frame.";
RL Mol. Gen. Genet. 204:126-132 (1986).
RN [2]
RP SEQUENCE OF 1-131 FROM N.A.
RX STRAIN=K12 / MGI655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119 (1995).
RN [3]
RP SEQUENCE.
RX STRAIN=K;
RX MEDLINE=77225229; PubMed=328274;
RA Hitz H., Schaefer D., Wittmann-liebold B.;
RT "Determination of the complete amino-acid sequence of protein S6 from
RT the wild-type and a mutant of Escherichia coli.";
RL Eur. J. Biochem. 75:497-512 (1977).
RN [4]
RP SEQUENCE OF 1-20.
RX STRAIN=K12;
RX MEDLINE=99085675; PubMed=9868784;
RA Wasinger V.C., Humphrey-Smith I.;
RT "Small genes/gene-products in Escherichia coli K-12.";
RL FEMS Microbiol. Lett. 169:375-382 (1998).
RN [5]
RP SEQUENCE OF 1-11.
RX STRAIN=K12 / W3110;
RA Paquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
RA Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,
RA Hochstrasser D.F.;
RT Submitted (SEP-1994) to the SWISS-PROT data bank.
RN [6]
RP SEQUENCE OF 1-12.
RX STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313 (1997).
RN [7]
RP SEQUENCE OF 1-4.
RX STRAIN=K12 / W3110;
RX MEDLINE=98263247; PubMed=9600841;
RA Wilkins M.R., Gasteiger E., Tonella L., Ou K., Tyler M.,
RA Sanchez J.-C., Gooley A.A., Walsh B.J., Bairoch A., Appel R.D.,
RA Williams K.L., Hochstrasser D.F.;
RT "Protein identification with N and C-terminal sequence tags in
RT proteome projects.";
RL J. Mol. Biol. 278:599-608 (1998).
RN [8]
RP MASS SPECTROMETRY.
RX STRAIN=K12 / ATCC 25404;
RX MEDLINE=99196679; PubMed=10094780;
RA Arnold R.J., Reilly J.P.;
RT "Observation of Escherichia coli ribosomal proteins and their
RT posttranslational modifications by mass spectrometry.";

```

```
RL Anal. Biochem. 269:105-112(1999).
RN [9]
RP 3D-STRUCTURE MODELING.
RX MEDLINE=22239879; PubMed=12244297;
RA Tung C.-S., Joseph S., Sanbonmatsu K.Y.;
RT "All-atom homology model of the Escherichia coli 30S ribosomal
   subunit.";
RL Nat. Struct. Biol. 9:750-755(2002).
CC -!- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA.
CC -!- PTM: 5 DIFFERENT FORMS OF THE PROTEIN, VARYING ONLY IN THE NUMBER
CC OF CARBOXYL-TERMINAL GLUTAMIC ACID RESIDUES, WERE ISOLATED. THE
CC SEQUENCE SHOWN IS FORM S6-6, WHICH IS THE LONGEST. THE FIRST 2 GLU
CC ARE ENCODED BY THE RPSF GENE, THE OTHER GLU ARE ADDED POST-
CC TRANSLATIONALLY BY THE RIMK ENZYME.
CC -!- MASS SPECTROMETRY: MW=15187.2; METHOD=MALDI; RANGE=1-131.
CC -!- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04022; CAA27652.1; -.
DR EMBL; U14003; AAA97096.1; -.
DR EMBL; AE000491; AAC77157.1; -.
DR PDB; 1M5G; 09-JUL-02.
DR SWISS-2DPAGE; P02358; COLI.
DR ECO2DBASE; D014.7; 6TH EDITION.
DR ECO2DBASE; C014.8; 6TH EDITION.
DR ECO2DBASE; C015.3; 6TH EDITION.
DR EcoGene; EGI0905; rpsf.
DR HAMAP; MF 00360; -.
DR InterPro; IPR000529; Ribosomal S6.
DR Pfam; PF01250; Ribosomal S6; 1.
DR ProDom; PD003809; Ribosomal_S6; 1.
DR TIGRFAMs; TIGR00166; S6; 1.
DR PROSITE; PS01048; RIBOSOMAL_S6; 1.
DR Ribosomal protein; rRNA-binding; 3D-structure; Complete proteome.
FT CHAIN 1 132 30S RIBOSOMAL PROTEIN S6, NON-MODIFIED
FT ISOFORM
FT CHAIN 1 135 30S RIBOSOMAL PROTEIN S6, FULLY MODIFIED
FT ISOFORM.
FT CONFLICT 14 14 Q -> T (IN REF. 4).
FT CONFLICT 20 20 G -> A (IN REF. 4).
SQ SEQUENCE 135 AA; 15703 MW; F4CC629711C1FD0E CRC64;

Query Match 68.0%; Score 34; DB 1; Length 135;
Best Local Similarity 62.5%; Pred. No. 9.5;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHYLV 9
Db 54 LHKAHYVL 61
```

Search completed: December 16, 2003, 14:15:16
Job time : 6.66667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:09:45 ; Search time 25 Seconds
(without alignments)
92.899 Million cell updates/sec

Title: US-09-870-089b-11
Perfect score: 50
Sequence: 1 FLHKVHYLV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	80.0	729	13 Q8AYJ5	Q8ayj5 brachydanio
2	39	78.0	1885	3 Q8TGV2	Q8tgv2 exophiala d
3	37	74.0	152	6 Q9GKX9	Q9gkx9 sus scrofa
4	37	74.0	364	6 Q77563	Q77563 sus scrofa
5	37	74.0	439	5 Q8T2D3	Q8t2d3 dictyosteli
6	36	72.0	187	11 Q8K4E4	Q8k4e4 mus musculus
7	36	72.0	230	16 Q8XK08	Q8xk08 clostridium
8	36	72.0	321	10 Q94K11	Q94k11 arabidopsis
9	36	72.0	333	11 Q8K4F0	Q8k4f0 mus musculus
10	36	72.0	378	10 Q23227	Q23227 arabidopsis
11	36	72.0	454	16 Q92CR5	Q92cr5 listeria in
12	36	72.0	454	16 Q87X77	Q87x77 listeria mo
13	36	72.0	670	16 Q9ZCM1	Q9zcm1 rickettsia
14	36	72.0	702	5 Q8IJJ8	Q8ijj8 plasmodium
15	36	72.0	716	5 Q9VHM2	Q9vhm2 drosophila
16	36	72.0	942	5 Q960B2	Q960b2 drosophila

17	36	72.0	1078	5 Q81077	Q8i077 drosophila
18	36	72.0	1084	5 Q9VAM3	Q9vam3 drosophila
19	36	72.0	1673	5 Q77332	Q77332 plasmodium
20	35	70.0	92	10 Q8S9C2	Q8s9c2 brassica ol
21	35	70.0	92	10 Q8RVY8	Q8rvy8 brassica ca
22	35	70.0	92	10 Q8S9C1	Q8s9c1 brassica ol
23	35	70.0	92	10 Q8RVZ0	Q8rvz0 brassica ca
24	35	70.0	92	10 Q8RVY9	Q8rvy9 brassica ca
25	35	70.0	182	16 Q8CND4	Q8cnd4 staphylococ
26	35	70.0	242	4 Q8TJD7	Q8tdj7 homo sapien
27	35	70.0	347	5 Q01535	Q01535 caenorhabdi
28	35	70.0	378	6 Q9TTY4	Q9tty4 canis famil
29	35	70.0	457	4 Q8TDJ8	Q8tdj8 homo sapien
30	35	70.0	650	5 Q44236	Q44236 ciona intes
31	35	70.0	662	16 Q92GP2	Q92gp2 rickettsia
32	35	70.0	763	4 Q8TJD9	Q8tdj9 homo sapien
33	35	70.0	929	4 Q8NGZ7	Q8ngz7 homo sapien
34	35	70.0	1645	5 Q27448	Q27448 babesia bov
35	34	68.0	58	16 Q92HR4	Q92hr4 rickettsia
36	34	68.0	123	16 Q8DCL6	Q8dcl6 vibrio vuln
37	34	68.0	131	16 Q8EAH2	Q8eah2 shewanella
38	34	68.0	136	10 Q94CT9	Q94ct9 oryza sativ
39	34	68.0	151	16 Q8R9B9	Q8r9b9 thermoanaer
40	34	68.0	209	4 Q8BT16	Q8bt16 homo sapien
41	34	68.0	209	11 Q9CWL2	Q9cwl2 mus musculu
42	34	68.0	227	4 Q9UQ17	Q9uq17 homo sapien
43	34	68.0	261	17 Q9Y977	Q9y977 aeropyrum p
44	34	68.0	303	16 Q8Z0Y4	Q8z0y4 salmonella
45	34	68.0	338	4 Q8N798	Q8n798 homo sapien

ALIGNMENTS

RESULT 1

Q8AYJ5 PRELIMINARY; PRT; 729 AA.
AC Q8AYJ5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Muskelein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Adams J.C., Zaromytidou A.-I.;
RT "zebrafish muskelein from cDNA."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF418017; AAN32664.1; -
SQ SEQUENCE 729 AA; 84376 MW; DFE3672F5D0AE16F CRC64;

Query Match 80.0%; Score 40; DB 13; Length 729;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYL 8
Db 588 LHKVHYL 594

RESULT 2

Q8TGV2 PRELIMINARY; PRT; 1885 AA.
AC Q8TGV2;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Chitin synthase V.
GN CHS5.

```

OS Exophiala dermatitidis (Wangiella dermatitidis).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Chaetothyriomycetes;
OC Chaetothyriales; Herpotrichiellaceae; mitosporic Herpotrichiellaceae;
OC Exophiala.
OX NCBI_TaxID=5970;
RN [1]
RN SEQUENCE FROM N.A.
RA Liu H., Szanislo P.J.;
RT "Cloning and Characterization of a Class V Chitin Synthase Gene
RT (WdCHS5) of Wangiella (Exophiala) dermatitidis, A Model Dematiaceous
RT Pathogen of Humans.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF469116; AAL79830.1; -.
DR InterPro; IPR001199; Cyt_B5.
DR InterPro; IPR004835; Fungi_chitin_syn.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF03142; Chitin_synth_2; 1.
DR Pfam; PF00173; heme_1; 2.
DR Pfam; PF00063; myosin_head; 1.
DR ProDom; PD000612; Cyt_B5; 1.
DR SMART; SM00242; MYSG; 1.
DR PROSITE; PS0255; CYTOCHROME_B5_2; 1.
SQ SEQUENCE 1885 AA; 208884 MW; 43F37BFABE5062BC CRC64;

Query Match 78.0%; Score 39; DB 3; Length 1885;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHY 7
DB 1370 FLHRVHY 1376

RESULT 3
Q9GKX9 PRELIMINARY; PRT; 152 AA.
ID Q9GKX9
AC Q9GKX9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE A transferase (Fragment).
GN ABO.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RN SEQUENCE FROM N.A.
RA Yamamoto M., Kominato Y., Noda R., Saitou N., Yamamoto F.;
RT "Mammalian ABO genes: Murine AB gene encodes a glycosyltransferase
RT with both A and B transferase activity and porcine O gene is a null
RT allele due to a major chromosomal aberration.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB041040; BAB20561.1; -.
DR InterPro; IPR005076; Glyco_trans_6.
DR Pfam; PF03414; Glyco_transf_6; 1.
DR Transferrase.
KW NON TER 1
FT NON TER 152
FT NON TER 152
SQ SEQUENCE 152 AA; 17442 MW; BCBD22A1E3DEF7EC CRC64;

Query Match 74.0%; Score 37; DB 6; Length 152;
Best Local Similarity 77.8%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHYLV 9
DB 54 FLHEVDYLV 62

RESULT 4
O77563 PRELIMINARY; PRT; 364 AA.
ID O77563

```

```

AC O77563;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative blood group A transferase.
GN A.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RN SEQUENCE FROM N.A.
RA Meijerink E., Voegeli P., Stranzinger G.;
RT "The putative porcine blood group A transferase cDNA.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF050177; AAC68840.1; -.
DR InterPro; IPR005076; Glyco_trans_6.
DR Pfam; PF03414; Glyco_transf_6; 1.
DR Transferrase.
KW SEQUENCE 364 AA; 42258 MW; 3D51850110EE75EA CRC64;

Query Match 74.0%; Score 37; DB 6; Length 364;
Best Local Similarity 77.8%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHYLV 9
DB 210 FLHEVDYLV 218

RESULT 5
Q8T2D3 PRELIMINARY; PRT; 439 AA.
ID Q8T2D3
AC Q8T2D3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 49.6 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gieckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115613; AAM10762.1; -.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000648; Oxysterol_BP.
DR Pfam; PF01237; Oxysterol_BP; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 49590 MW; 65F0C94E40E4AAD9 CRC64;

Query Match 74.0%; Score 37; DB 5; Length 439;
Best Local Similarity 55.6%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLHKVHYLV 9
DB 380 FLHKPHYFI 388

RESULT 6
Q8K4E4 PRELIMINARY; PRT; 187 AA.
ID Q8K4E4
AC Q8K4E4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```

```
DE Platelet and T cell activation antigen 1 isoform 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=97413766; PubMed=9268302;
RA Sherrington P.D., Scott J.L., Jin B., Simmons D., Dorahy S.J.,
RA Lloyd J., Brien J.H., Aebersold R.H., Adamson J., Zuzel M.,
RA Burns G.F.;
RT "TLISA1 (PTA1) activation antigen implicated in T cell differentiation
RT and platelet activation is a member of the immunoglobulin superfamily
RT exhibiting distinctive regulation of expression.";
RL J. Biol. Chem. 272:21735-21744(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=20112343; PubMed=10647817;
RA Tian F., Li D., Xia H., Liu X., Jia W., Sun C., Sun K., Jin B.;
RT "Isolation of cDNAs encoding gibbon and monkey platelet and T cell
RT activation antigen 1 (PTA1).";
RL DNA Seq. 10:155-161(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX STRAIN=BALB/C;
RA Jin B., Zhang X., Li D., Ouyang W., Jia W., Chen L., Xie X., Ning S.,
RA Zhang Y.;
RT "Gene cloning and characterization of mouse platelet and T cell
RT activation antigen 1 (PTA1/CD226).";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF421198; AAN04296.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; ig; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR SEQUENCE 187 AA; 21393 MW; 70D99BE21927ACE5 CRC64;

Query Match 72.0%; Score 36; DB 11; Length 187;
Best Local Similarity 62.5%; Pred. No. 43;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLKHVHYL 8
Db 76 YLHRVHFL 83

RESULT 7
Q8XKM8 PRELIMINARY; PRT; 250 AA.
AC Q8XKM8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein CPE1366.
GN CPE1366.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RL EMBL; AP003190; BAB81072.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 250 AA; 29634 MW; 60F81BD7DF6C973 CRC64;
```

```
Query Match 72.0%; Score 36; DB 16; Length 250;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHYLV 9
Db 22 LHYVHYLI 29

RESULT 8
Q94K11 PRELIMINARY; PRT; 321 AA.
AC Q94K11;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Unknown protein (EC 3.3.2.3) (Epoxide hydrolase).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lam B., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BIOTRANSFORMATION ENZYME THAT CATALYZES THE HYDROLYSIS
CC OF EPOXIDES (ALKENE OXIDES, OXIRANES) AND ARENE OXIDES TO LESS
CC REACTIVE AND MORE WATER SOLUBLE DIHYDRODIOLS BY THE TRANS ADDITION
CC OF WATER (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: AN EPOXIDE + H(2)O = A GLYCOL.
DR EMBL; AF370481; AAK43858.1; -.
DR InterPro; IPR000073; A/b hydrolase.
DR InterPro; IPR003089; AB hydrolase.
DR InterPro; IPR000639; Epox hydrolase.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00561; abhydrolase; 1.
DR PRINTS; PR00111; ABHYDROLASE.
DR PRINTS; PR00412; EPOXYHYDROLASE.
KW Aromatic hydrocarbons catabolism; Detoxification; Hydrolase.
SQ SEQUENCE 321 AA; 35777 MW; C1031E942DD5F34 CRC64;
```

```
Query Match 72.0%; Score 36; DB 10; Length 321;
Best Local Similarity 71.4%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHYLV 9
Db 32 HKIHYVV 38

RESULT 9
Q8K4F0 PRELIMINARY; PRT; 333 AA.
AC Q8K4F0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Platelet and T cell activation antigen 1.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Thymus;
RA Jin B., Zhang X., Li D., Ouyang W., Jia W., Chen L., Xie X., Ning S.,
```

RA Zhang Y.;
 RT "Gene cloning and characterization of mouse platelet and T cell
 RT activation antigen 1 (PTAI/CD226).";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF416980; AAN04295.1; -.

DR InterPro; IPR003599; Ig.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003006; Ig_MHC.

DR Pfam; PF00047; Ig; 2.

DR SMART; SM00409; Ig; 2.

DR PROSITE; PS00835; IG_LIKE; 2.

SQ SEQUENCE 333 AA; 38063 MW; ACDE524D0F475C97 CRC64;

Query Match 72.0%; Score 36; DB 11; Length 333;

Best Local Similarity 62.5%; Pred. No. 75;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLHKVHYL 8

:|:|:|

76 YLHRVHFL 83

RESULT 10

O23227 O23227 PRELIMINARY; PRT; 378 AA.

AC O23227;

DT 01-JAN-1998 (TRENBLrel. 05, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Hypothetical protein (EC 3.3.2.3) (Epoxide hydrolase).

GN C7A10.830 OR AT4G36530.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

[1]

RP SEQUENCE FROM N.A.

RA Bevan M., Terryn N., Vos P., Heijnen L., Mewes H.W., Schueller C.,

RA Chludzisz N.;

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

[2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[3]

RP SEQUENCE FROM N.A.

RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,

RA Feldmann K.A., Flavell R.B., White O., Salzberg S.B.;

RT "Full-length messenger RNA sequences greatly improve genome

RT annotation.";

RL Genome Biol. 0:0-0(2002).

[4]

RP SEQUENCE FROM N.A.

RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,

RA Feldmann K.;

RT "Full-length cDNA from Arabidopsis thaliana.";

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: BIOTRANSFORMATION ENZYME THAT CATALYZES THE HYDROLYSIS

CC OF EPOXIDES (ALKENE OXIDES, OXIRANES) AND ARENE OXIDES TO LESS

CC REACTIVE AND MORE WATER SOLUBLE DIHYDRODIOLS BY THE TRANS ADDITION

CC OF WATER (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: AN EPOXIDE + H(2)O = A GLYCOL.

DR EMBL; Z99708; CAB16840.1; -.

DR EMBL; AL161589; CAB80319.1; -.

DR EMBL; AY087588; AAM65130.1; -.

DR InterPro; IPR000073; A/B_hydrolase.

DR InterPro; IPR003089; AB_hydrolase.

DR InterPro; IPR000639; Epox_hydrolase.

DR InterPro; IPR000379; Ser_estrns_site.

DR Pfam; PF00561; abhydrolase; 1.

DR PRINTS; PR00111; ABHYDROLASE

DR PRINTS; PR00412; EPOXYDOLASE.

KW Hypothetical protein; Aromatic hydrocarbons catabolism;
 KW Detoxification; Hydrolase.

SQ SEQUENCE 378 AA; 41906 MW; 8D1FB39D5410E818 CRC64;

Query Match 72.0%; Score 36; DB 10; Length 378;

Best Local Similarity 71.4%; Pred. No. 85;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HKVHYLV 9

:|:|:|

89 HKIHYV 95

RESULT 11

Q92CR5 Q92CR5 PRELIMINARY; PRT; 454 AA.

AC Q92CR5;

DT 01-DEC-2001 (TRENBLrel. 19, Created)

DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)

DE Hypothetical protein lin1106.

GN LIN106

OS Listeria innocua.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI_TaxID=1642;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLIP 11262 / Serovar 6a;

RX PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

RA Baguero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,

RA Charbit A., Cherouani F., Couve E., de Daruvar A., Dehoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,

RA Enian K.-D., Fejhi H., Garcia-del Portillo F., Garrido P.,

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,

RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,

RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordmel B., Rose M., Schluter T., Simoes N., Tierrez A.,

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;

RT "Comparative genomics of Listeria species.";

RL Science 294:849-852(2001).

DR EMBL; AL596167; CAC96337.1; -.

DR Listifist; LIN01106; -.

DR InterPro; IPR001450; 4Fe4S_ferredoxin.

DR Pfam; PF01512; Complex1_51K; 1.

DR Pfam; PF00037; fer4; 1.

DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 454 AA; 49918 MW; 068B13889213B239 CRC64;

Query Match 72.0%; Score 36; DB 16; Length 454;

Best Local Similarity 71.4%; Pred. No. 1e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLHKVHY 7

:|:|:|

239 YLKLHY 245

RESULT 12

Q8Y7X7 Q8Y7X7 PRELIMINARY; PRT; 454 AA.

AC Q8Y7X7;

DT 01-MAR-2002 (TRENBLrel. 20, Created)

DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)

DE Hypothetical protein lmo1142.

GN LMO1142

OS Listeria monocytogenes.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI_TaxID=1639;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=EGD-e / Serovar 1/2a;
RX  MEDLINE=21537279; PubMed=11679669;
RA  Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amand A.,
RA  Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA  Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA  Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA  Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA  Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA  Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,
RA  Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA  Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA  Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA  Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;
RT  "Comparative genomics of Listeria species.";
RL  Science 294:849-852(2001).
DR  EMBL; AL591978; CAC99220.1; -.
DR  ListList; LMO01142; -.
DR  InterPro; IPR001450; 4Pe4S_ferredoxin.
DR  InterPro; IPR001949; Complex1_51K.
DR  Pfam; PF01512; Complex1_51K; 1.
DR  Pfam; PF00037; fer4; 1.
DR  PROSITE; PS00198; 4Pe4S_FERREDOXIN; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 454 AA; 49891 MW; 006AD8F5C732B44B CRC64;

Query Match 72.0%; Score 36; DB 16; Length 454;
Best Local Similarity 71.4%; Pred. No. 1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  1 FLHKVHY 7
DB  239 YLKLHY 245

RESULT 13
Q9ZCM1 PRELIMINARY; PRT; 670 AA.
AC Q9ZCM1;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cytochrome C-type biogenesis protein CCMP (CCMP).
GN Rpt03.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsia; Rickettsia.
OX NCBI_TaxID=782;
RN [1]_TaxID=782;
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
DR EMBL; AJ235273; CAA15139.1; -.
DR InterPro; IPR002541; CytC_asm.
DR InterPro; IPR003567; CytC_biol.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF01578; CytC_asm; 1.
DR PRINTS; PR01410; CCBI0GENESIS.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
KW Complete proteome.
SQ SEQUENCE 670 AA; 76847 MW; 2703A1AC44F9BBD CRC64;

Query Match 72.0%; Score 36; DB 16; Length 670;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY  2 LHKVHYL 9
DB  500 LHNHYLL 507

RESULT 14
Q8IUJ8 PRELIMINARY; PRT; 702 AA.
AC Q8IUJ8;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF10_0290.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_TaxID=7227;
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shalim S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014833; AAN35487.1; -.
KW Hypothetical protein.
SQ SEQUENCE 702 AA; 82907 MW; 37D1DB57A12AE6B CRC64;

Query Match 72.0%; Score 36; DB 5; Length 702;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY  1 FLHKVHYL 8
DB  526 FOHRIHYL 533

RESULT 15
Q9VHM2 PRELIMINARY; PRT; 716 AA.
AC Q9VHM2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE CG8202 protein.
GN CG8202.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]_TaxID=7227;
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

```

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodge T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003680; AAF54281.2; -;
DR FlyBase; FBgn0037622; CG8202;
SQ SEQUENCE 716 AA; 82149 MW; 4DB233CE393C5704 CRC64;

Query Match 72.0%; Score 36; DB 5; Length 716;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHY 7
Db 622 FIHKVHF 628
Search completed: December 16, 2003, 14:18:03
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:11:00 ; Search time 12.3333 Seconds
(without alignments)
30.875 Million cell updates/sec

Title: US-09-870-089B-13

Perfect score: 55

Sequence: 1 KHFKPHGFS 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTCTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	74.5	2391	2	US-08-446-855A-2
2	41	74.5	2391	3	US-09-150-741-2
3	37	67.3	17	4	US-09-425-679E-8
4	37	67.3	110	4	US-09-425-679E-10
5	37	67.3	508	4	US-09-107-532A-5331
6	37	67.3	1288	1	US-07-727-814B-2
7	37	67.3	1288	1	US-08-258-614-2
8	35	63.6	512	2	US-08-828-596-2
9	35	63.6	512	3	US-09-360-197-6
10	35	63.6	908	4	US-09-635-872A-15
11	35	63.6	908	4	US-09-636-077A-15
12	34	61.8	256	4	US-09-252-991A-19828
13	34	61.8	363	4	US-09-252-991A-25132
14	34	61.8	405	3	US-09-347-801-20
15	34	61.8	462	4	US-09-328-352-6483
16	34	61.8	534	4	US-09-328-352-6713
17	34	61.8	1463	1	US-08-220-603A-11
18	33	60.0	255	4	US-09-134-001C-3680
19	33	60.0	299	2	US-08-440-845D-24
20	33	60.0	364	4	US-09-328-352-4956
21	33	60.0	382	1	US-08-464-523B-30
22	33	60.0	382	2	US-08-440-845D-2
23	33	60.0	382	3	US-08-868-458-2
24	33	60.0	490	4	US-09-252-991A-30355
25	33	60.0	786	4	US-09-012-710-11
26	33	60.0	786	4	US-09-556-273-11
27	33	60.0	1452	4	US-09-252-991A-18118

28	32	58.2	58	4	US-09-732-210-480	Sequence 480, Appl
29	32	58.2	62	4	US-09-732-210-483	Sequence 483, Appl
30	32	58.2	88	4	US-09-732-210-1179	Sequence 1179, Appl
31	32	58.2	183	1	US-08-157-005-6	Sequence 6, Appli
32	32	58.2	183	2	US-08-799-464A-20	Sequence 20, Appli
33	32	58.2	183	3	US-08-747-863-6	Sequence 6, Appli
34	32	58.2	183	3	US-08-478-316-31	Sequence 31, Appli
35	32	58.2	183	4	US-09-019-793A-31	Sequence 31, Appli
36	32	58.2	183	4	US-09-565-864-6	Sequence 6, Appli
37	32	58.2	183	4	US-09-328-352-6035	Sequence 6035, Ap
38	32	58.2	183	5	PCT-US95-09927-20	Sequence 20, Appl
39	32	58.2	183	5	PCT-US95-10904-77	Sequence 77, Appl
40	32	58.2	208	4	US-09-252-991A-20307	Sequence 20307, A
41	32	58.2	333	4	US-09-107-532A-5293	Sequence 5293, Ap
42	32	58.2	337	4	US-09-252-991A-26757	Sequence 26757, A
43	32	58.2	371	4	US-09-252-991A-19630	Sequence 19630, A
44	32	58.2	399	4	US-09-252-991A-18540	Sequence 18540, A
45	32	58.2	412	4	US-09-252-991A-26961	Sequence 26961, A

ALIGNMENTS

RESULT 1
US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-855A-2

Query Match 74.5%; Score 41; DB 2; Length 2391;
Best Local Similarity 77.8%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9

Db 1168 KYFKHGFS 1176

RESULT 2
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

Query Match 74.5%; Score 41; DB 3; Length 2391;
Best Local Similarity 77.8%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHKPHGFS 9
Db 1168 KYFKHGFS 1176

RESULT 3
US-09-425-679E-8
; Sequence 8, Application US/09425679E
; Patent No. 6509444
; GENERAL INFORMATION:
; APPLICANT: McDonald, Thomas L.
; TITLE OF INVENTION: Serum Amyloid A Isoform from Colostrum
; FILE REFERENCE: UNMC 63142
; CURRENT APPLICATION NUMBER: US/09/425,679E
; CURRENT FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: US 60/218,482
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/218,611
; PRIOR FILING DATE: 2000-07-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Equus caballus
US-09-425-679E-8

Query Match 67.3%; Score 37; DB 4; Length 17;
Best Local Similarity 83.3%; Pred. No. 1;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HFKPHG 7
Db 7 HFRPHG 12

RESULT 4
US-09-425-679E-10

; Sequence 10, Application US/09425679E
; Patent No. 6509444
; GENERAL INFORMATION:
; APPLICANT: McDonald, Thomas L.
; TITLE OF INVENTION: Serum Amyloid A Isoform from Colostrum
; FILE REFERENCE: UNMC 63142
; CURRENT APPLICATION NUMBER: US/09/425,679E
; CURRENT FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: US 60/218,482
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/218,611
; PRIOR FILING DATE: 2000-07-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Equus caballus
US-09-425-679E-10

Query Match 67.3%; Score 37; DB 4; Length 110;
Best Local Similarity 83.3%; Pred. No. 7.5;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HFKPHG 7
Db 100 HFRPHG 105

RESULT 5
US-09-107-532A-5331
; Sequence 5331, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: STC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5331:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES

US-09-107-532A-5331
; Sequence 5331, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: STC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5331:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES


```
/ ORIGINAL SOURCE:
/ ORGANISM: Enterococcus faecium
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...508
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5331:
US-09-107-532A-5331

Query Match          67.3%; Score 37; DB 4; Length 508;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 HKPKPGF 8
Db      293 HKKPHGF 299

RESULT 6
US-07-727-814B-2
/ Sequence 2, Application US/07727814B
/ Patent No. 5213969
/ GENERAL INFORMATION:
/ APPLICANT: SCHUMACHER, Gunther
/ APPLICANT: BURTSCHER, Helmut
/ APPLICANT: MOLLER, Hans
/ TITLE OF INVENTION: CLONED N-METHYLHYDANTOINASE
/ NUMBER OF SEQUENCES: 6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Armstrong, Nikaido, Marmelstein, Kubovcik &
/ ADDRESSEE: Murray
/ STREET: 1725 K Street, N.W., Suite 1000
/ CITY: Washington D.C.
/ COUNTRY: United States of America
/ ZIP: 20006
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/727,814B
/ FILING DATE: 19910708
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: DE P 4021571.7
/ FILING DATE: 06-JUL-1990
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Murray, Robert B.
/ REGISTRATION NUMBER: 22,980
/ REFERENCE/DOCKET NUMBER: 911014
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 659-2930
/ TELEX: 440142
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1288 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-07-727-814B-2

Query Match          67.3%; Score 37; DB 1; Length 1288;
Best Local Similarity 66.7%; Pred. No. 1.le+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 HKPKPHGFS 9
Db      97 RHKKPHNFS 105

RESULT 7
US-08-258-614-2
/ Sequence 2, Application US/08258614
/ Patent No. 5432070
/ GENERAL INFORMATION:
/ APPLICANT: SCHUMACHER, Gunther
/ APPLICANT: BURTSCHER, Helmut
/ APPLICANT: MOLLER, Hans
/ TITLE OF INVENTION: CLONED N-METHYLHYDANTOINASE
/ NUMBER OF SEQUENCES: 6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Nikaido, Marmelstein, Murray & Oram
/ STREET: 655 Fifteenth Street N.W. Suite 330
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20005-5701
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/258,614
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/021,856
/ FILING DATE: 24-FEB-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: DE P 40 21 571.7
/ FILING DATE: 06-JUL-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/727,814
/ FILING DATE: 08-JUL-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kitts, Monica C.
/ REGISTRATION NUMBER: 36,105
/ REFERENCE/DOCKET NUMBER: P564-3007
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202)638-5000
/ TELEFAX: (202)638-4810
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1288 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-258-614-2

Query Match          67.3%; Score 37; DB 1; Length 1288;
Best Local Similarity 66.7%; Pred. No. 1.le+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 HKPKPHGFS 9
Db      97 RHKKPHNFS 105

RESULT 8
US-08-828-596-2
/ Sequence 2, Application US/08828596
/ Patent No. 5892018
/ GENERAL INFORMATION:
/ APPLICANT: Welsh, Michael J.
/ APPLICANT: Price, Margaret P.
/ TITLE OF INVENTION: No. 5892018el Brain Sodium Channel Protein Family
/ TITLE OF INVENTION: and DNA Sequences Encoding Same
/ NUMBER OF SEQUENCES: 5
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Zarley, McKee, Thomte, Voorhees & Sease
/ STREET: 801 Grand Suite 3200
/ CITY: Des Moines
/ STATE: Iowa
/ COUNTRY: United States
```

; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,596
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/626,838
; FILING DATE: 02-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nebel, Heidi S.
; REGISTRATION NUMBER: 37,719
; REFERENCE/DOCKET NUMBER: uirf n6-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-288-3667
; TELEFAX: 515-288-1338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-828-596-2

Query Match 63.6%; Score 35; DB 2; Length 512;
Best Local Similarity 66.7%; Pred. No. 93;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KHKPKHGFS 9
||:||||
Db 144 KHKPKQFS 152

RESULT 9
US-09-360-197-6
; Sequence 6, Application US/09360197
; Patent No. 6287859
; GENERAL INFORMATION:
; APPLICANT: Bassilana, Frederic
; APPLICANT: Lazdunski, Michel
; APPLICANT: Waldmann, Rainer
; APPLICANT: Deweille, Jan R.
; TITLE OF INVENTION: Human and Rat Families of Neuronal Acid-Sensitive
; TITLE OF INVENTION: Cationic Channels, Their Cloning and Applications
; FILE REFERENCE: 989,6706P
; CURRENT APPLICATION NUMBER: US/09/360,197
; CURRENT FILING DATE: 1997-07-23
; PRIOR FILING DATE: 09/129,758
; PRIOR FILING DATE: 1998-08-05
; PRIOR APPLICATION NUMBER: 60/095,408
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 6
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-360-197-6

Query Match 63.6%; Score 35; DB 3; Length 512;
Best Local Similarity 66.7%; Pred. No. 93;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KHKPKHGFS 9
||:||||
Db 144 KHKPKQFS 152

RESULT 10

US-09-635-872A-15
; Sequence 15, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613USO
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 15
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-635-872A-15

Query Match 63.6%; Score 35; DB 4; Length 908;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HFKPHGFS 9
|:||||:
Db 771 HRRPHGFA 778

RESULT 11
US-09-636-077A-15
; Sequence 15, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612USO
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 15
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-077A-15

Query Match 63.6%; Score 35; DB 4; Length 908;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HFKPHGFS 9
|:||||:
Db 771 HRRPHGFA 778

RESULT 12
US-09-252-991A-19828
; Sequence 19828, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```
; SEQ ID NO 19828
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19828

Query Match          61.8%; Score 34; DB 4; Length 256;
Best Local Similarity 71.4%; Pred. No. 67;
Matches 5; Conservative 1; Mismatches 0; Indels 1; Gaps 0;

QY 1 KHKFKPHG 7
Db 113 QHKFKPHG 119

RESULT 13
US-09-252-991A-25132
; Sequence 25132, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25132
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25132

Query Match          61.8%; Score 34; DB 4; Length 363;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HFKPH 6
Db 218 HFKPH 222

RESULT 14
US-09-347-801-20
; Sequence 20, Application US/09347801
; Patent No. 6262345
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; TITLE OF INVENTION: Plant Protein Kinases
; FILE REFERENCE: BB-1171
; CURRENT APPLICATION NUMBER: US/09/347,801
; PRIOR FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,438
; EARLIER FILING DATE: July 10, 1998
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-347-801-20

Query Match          61.8%; Score 34; DB 3; Length 405;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HFKPH 6
Db 113 QHKFKPHG 119
```

```
Db 374 HFKPH 378

RESULT 15
US-09-328-352-6483
; Sequence 6483, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6483
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6483

Query Match          61.8%; Score 34; DB 4; Length 462;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 FPKHGFS 9
Db 202 FPKHGFS 208

Search completed: December 16, 2003, 14:20:42
Job time : 13.3333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:10:15 ; Search time 10.3333 Seconds
(without alignments)
83.760 Million cell updates/sec

Title: US-09-870-089b-13

Perfect score: 55

Sequence: 1 KHFXPHGFS 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	DB ID	Description
1	55	100.0	351	2 A45377	transcription factor
2	49	89.1	381	2 S26812	transcription factor
3	43	78.2	348	1 H64657	probable cinnamyl-zinc-dependent alc
4	43	78.2	350	2 A71857	hypothetical prote
5	41	74.5	191	2 D70485	carbamoyl-phosphat conserved hypothet
6	41	74.5	2391	2 T18410	hypothetical prote
7	39	70.9	375	2 A00095	hypothetical prote
8	39	70.9	633	2 T19189	hypothetical prote
9	39	70.9	639	2 G88839	protein C10C6.5 [i
10	38	69.1	341	2 S45524	5'-phosphoribosyl-
11	37	67.3	110	2 A28573	serum amyloid A pr
12	37	67.3	310	2 T28848	hypothetical prote
13	37	67.3	411	2 T19195	hypothetical prote
14	37	67.3	642	2 A65600	hypothetical prote
15	37	67.3	668	2 T29884	hypothetical prote
16	37	67.3	754	2 G90484	beta-xylosidase [i
17	36	65.5	361	2 B81190	glycosyl hydrolase
18	36	65.5	361	2 H81913	probable beta-N-ac
19	36	65.5	381	2 S48049	cholecystokinin B
20	36	65.5	417	2 T26670	hypothetical prote
21	35	63.6	48	2 B90531	50S ribosomal prote
22	35	63.6	116	2 F83975	hypothetical prote
23	35	63.6	120	2 F69873	hypothetical prote
24	35	63.6	123	1 B64455	hypothetical prote
25	35	63.6	154	2 A36591	superoxide dismuta
26	35	63.6	244	1 H64806	ybgJ protein - Esc
27	35	63.6	244	2 B90721	probable lactam ut
28	35	63.6	244	2 H85571	probable lactam ut
29	35	63.6	244	2 AF0588	conserved hypothet

30	35	63.6	250	2 T10404	hypothetical prote
31	35	63.6	264	2 T51304	splicing factor RS
32	35	63.6	271	2 T47978	splicing factor RS
33	35	63.6	344	2 G82690	aspartate-B-semial
34	35	63.6	349	2 F85294	splicing factor At
35	35	63.6	350	2 T05797	splicing factor SR
36	35	63.6	367	2 B69111	conserved hypothet
37	35	63.6	368	2 S36576	E2 protein - human
38	35	63.6	375	2 T26701	hypothetical prote
39	35	63.6	384	2 I49528	hypothetical prote
40	35	63.6	401	2 D95423	RepC2 replication
41	35	63.6	592	2 E70455	sulfur oxidation p
42	35	63.6	629	2 T22066	hypothetical prote
43	35	63.6	868	1 A45391	mRNA guanylyltrans
44	34	61.8	45	2 A90081	50S ribosomal prot
45	34	61.8	138	2 F86715	hypothetical prote

ALIGNMENTS

RESULT 1

A45377
transcription factor ATF4 - human
N:Alternate names: activating transcription factor 4 (ATF-4); TAXREB67; transcription f.
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 02-Jun-2000
C:Accession: A45377; I56787; D34223
R:Karpinski, B.A.; Morle, G.D.; Huggenvik, J.; Uhler, M.D.; Leiden, J.M.
Proc. Natl. Acad. Sci. U.S.A. 89, 4820-4824, 1992
A:Title: Molecular cloning of human CREB-2: an ATF/CREB transcription factor that can n
A:Reference number: A45377; MUID:92279218; PMID:1534408
A:Accession: A45377
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-351 <KAR>
A:Cross-references: GB:M86842; NID:g181040; PIDN:AAA52071.1; PID:g181041
R:Tsujiimoto, A.; Wynnoya, H.; Morita, T.; Sato, T.; Shomotohno, K.
J. Virol. 65, 1420-1426, 1991
A:Title: Isolation of cDNA for DNA binding proteins which specifically bind to TAX-resp
A:Reference number: I56787; MUID:91140735; PMID:1847461
A:Accession: I56787
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-351 <TSU>
A:Cross-references: GB:D90209; NID:g220087; PIDN:BAA14234.1; PID:g220088
R:Hai, T.; Liu, F.; Coukos, W.J.; Green, M.R.
Genes Dev. 3, 2083-2090, 1989
A:Title: Transcription factor ATF cDNA clones: an extensive family of leucine zipper pr
A:Reference number: A91622; MUID:90185187; PMID:2516827
A:Accession: D34223
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 274-283, 'R', 285-289, 'R', 291-328, 'R', 330, 'K', 332-337, 'L', 339-341 <HA4>
C:Genetics:
A:Gene: GDB:ATF4; TXREB; TAXREB67; CREB-2
A:Cross-references: GDB:I32551
A:Map position: 22q12.3-22q13.1
C:Keywords: DNA binding; transcription factor

Query Match 100.0%; Score 55; DB 2; Length 351;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFXPHGFS 9

Db 42 KHFXPHGFS 50

RESULT 2

S26812
transcription factor ATF-4 - mouse
N:Alternate names: activating transcription factor 4; TAXREB67 homolog; transcription f;

C;Species: Mus musculus (house mouse)
 C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
 C;Accession: S26812; S18719; C46132; A58861; A47443
 R;Pruitt, S.C.
 submitted to the EMBL Data Library, August 1991
 A;Reference number: S26812
 A;Accession: S26812
 A;Molecule type: DNA
 A;Residues: 1-381 <PRU>
 A;Cross-references: EMBL:X61507; NID:g50049; PIDN:CAA43723.1; PID:g50050
 R;Mielnicki, L.M.; Pruitt, S.C.
 Nucleic Acids Res. 19, 6332, 1991
 A;Title: Isolation and nucleotide sequence of a murine cDNA homologous to human activator protein 1
 A;Reference number: S18719; MUID:92066493; PMID:1956797
 A;Accession: S18719
 A;Molecule type: DNA
 A;Residues: 1-57,59-288,'R',290-381 <MIE>
 A;Cross-references: EMBL:X61507
 R;Chevray, P.M.; Nathans, D.
 Proc. Natl. Acad. Sci. U.S.A. 89, 5789-5793, 1992
 A;Title: Protein interaction cloning in yeast: identification of mammalian proteins that interact with the E1A protein of human papilloma virus type 16
 A;Reference number: A46132; MUID:92335183; PMID:1631061
 A;Accession: C46132
 A;Status: nucleic acid sequence not shown; not compared with conceptual translation
 A;Molecule type: mRNA
 A;Residues: 246-381 <CHE1>
 A;Cross-references: GB:M94087; NID:g293723
 A;Experimental source: embryo
 A;Note: sequence extracted from NCBI backbone (NCBIP:108537)
 R;Chevray, P.M.; Nathans, D.
 submitted to GenBank, November 1994
 A;Reference number: A58861
 A;Accession: A58861
 A;Status: not compared with conceptual translation
 A;Molecule type: mRNA
 A;Residues: 'MSFLNSEVLADLMS',51-391 <CHRE2>
 A;Cross-references: GB:M94087; NID:g293723; PIDN:AAA53043.1; PID:g567226
 A;Experimental source: strain CD-1, embryo
 A;Note: The submission of the complete sequence to GenBank is mentioned in reference A46132
 R;Vallejo, M.; Ron, D.; Miller, C.P.; Habener, J.F.
 Proc. Natl. Acad. Sci. U.S.A. 90, 4679-4683, 1993
 A;Title: C/ATF, a member of the activating transcription factor family of DNA-binding proteins
 A;Reference number: A47443; MUID:93281642; PMID:8506317
 A;Accession: A47443
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 'MTMSFLNSEVLADLMS',51-376,'Q',378-381 <VAL>
 A;Cross-references: GB:L13791; NID:g293841; PIDN:AAA40476.1; PID:g293842
 C;Keywords: transcription factor

Query Match 89.1%; Score 49; DB 2; Length 381;
 Best Local Similarity 88.9%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHPKPHGFS 9
 ||| |||||
 Db 74 KHLKPHGFS 82

RESULT 3
 H64557
 probable cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) - Helicobacter pylori (strain 266)
 C;Species: Helicobacter pylori
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C;Accession: H64657
 R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
 Nature 388, 539-547, 1997
 A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A;Reference number: A64520; MUID:97394467; PMID:9252185
 A;Accession: H64657

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-348 <TCM>
 A;Cross-references: GB:AE000617; GB:AE000511; NID:g2314256; PIDN:AAD08150.1; PID:g231425
 C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C;Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
 F;27-333/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match 78.2%; Score 43; DB 1; Length 348;
 Best Local Similarity 87.5%; Pred. No. 2.5;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HFKPHGFS 9
 ||||| ||
 Db 16 HFKPHDFS 23

RESULT 4
 A71857
 zinc-dependent alcohol dehydrogenase - Helicobacter pylori (strain J99)
 C;Species: Helicobacter pylori
 A;Variety: strain J99
 C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 21-Jan-2000
 C;Accession: A71857
 R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen
 A;Reference number: A71800; MUID:99120557; PMID:9923682
 A;Accession: A71857
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-350 <ARN>
 A;Cross-references: GB:AE001531; GB:AE001439; NID:g4155617; PIDN:AAD06610.1; PID:g415562
 A;Experimental source: strain J99
 C;Genetics:
 A;Gene: jhp1030

Query Match 78.2%; Score 43; DB 2; Length 350;
 Best Local Similarity 87.5%; Pred. No. 2.5;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HFKPHGFS 9
 ||||| ||
 Db 16 HFKPHDFS 23

RESULT 5
 D70485
 hypothetical protein aq_2159 - Aquifex aeolicus
 C;Species: Aquifex aeolicus
 C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Jun-1999
 C;Accession: D70485
 R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
 V.
 Nature 392, 353-358, 1998
 A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A;Reference number: A70300; MUID:98196666; PMID:9537320
 A;Accession: D70485
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-191 <AQF>
 A;Cross-references: GB:AE000776; NID:g2984355; PIDN:AAC07886.1; PID:g2984367; GB:AE00065
 A;Experimental source: strain VF5
 C;Genetics:
 A;Gene: aq_2159

Query Match 74.5%; Score 41; DB 2; Length 191;
 Best Local Similarity 66.7%; Pred. No. 3.1;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
| : |||||
Db 119 KEYKPHGFN 127

RESULT 6
T18410
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18410
R:Flores, M.V.C.; O'Sullivan, W.J.; Stewart, T.S.
A:Description: Characterisation of the carbamoyl phosphate synthetase gene from Plasmodium falciparum
A:Reference number: Z18931
A:Accession: T18410
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2391 <FLO>
A:Cross-references: EMBL:L32150; NID:G476023; PID:G476024; PIDN:AAA29522.1
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology; B
arbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-phos
C:Keywords: ligase
F:96-2367/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>

Query Match 74.5%; Score 41; DB 2; Length 2391;
Best Local Similarity 77.8%; Pred. No. 41;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
| : |||||
Db 1168 KYFKKHGFS 1176

RESULT 7
AD0095
conserved hypothetical protein YPO0774 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AD0095
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0095
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89623.1; PID:g15978852; GSPDB:GN00175
C:Genetics:
A:Gene: YPO0774

Query Match 70.9%; Score 39; DB 2; Length 375;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HFKEPHGFS 9
| : |||||
Db 173 HFSPYGFS 180

RESULT 8
T19189
hypothetical protein C10C6.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T19189
R:White, S.
submitted to the EMBL Data Library, December 1996

A:Reference number: Z19086
A:Accession: T19189
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-633 <WIL>
A:Cross-references: EMBL:Z83217; PIDN:CAB05682.2; GSPDB:GN00022; CESP:C10C6.5
A:Experimental source: clone C10C6
C:Genetics:
A:Gene: CESP:C10C6.5
A:Map position: 4
A:Introns: 40/2; 70/2; 101/3; 132/3; 176/1; 263/3; 310/2; 379/3; 427/1; 522/3; 587/3
C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 70.9%; Score 39; DB 2; Length 633;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
| : |||||
Db 457 KHFRHGFA 465

RESULT 9
G88839
Protein C10C6.5 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: G88839
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.el
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an
A:Accession: G88839
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-639 <STO>
A:Cross-references: GB:chr_IV; PIDN:CAB05682.1; PID:g3874213; GSPDB:GN00022; CESP:C10C6
C:Genetics:
A:Gene: C10C6.5
A:Map position: 4
C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 70.9%; Score 39; DB 2; Length 639;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
| : |||||
Db 463 KHFRHGFA 471

RESULT 10
S45524
5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole synthetase - moth bean
C:Species: Vigna aconitifolia (moth bean)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 03-Mar-1995
C:Accession: S45524
R:Chapman, K.A.; Delauney, A.J.; Verma, D.P.S.
submitted to the EMBL Data Library, December 1992
A:Description: De novo purine biosynthesis in legume root nodules: cloning and sequenci
A:Reference number: S45524
A:Accession: S45524
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-341 <CHA>
A:Cross-references: EMBL:M92931

Query Match 69.1%; Score 38; DB 2; Length 341;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FKPHGF 8
|||||
Db 21 FKPHGF 26

RESULT 11

A28573
serum amyloid A protein - horse
N;Contains: amyloid protein AA
C;Species: Equus caballus (domestic horse)
C;Date: 19-Nov-1988 #sequence_revision 12-Feb-1993 #text_change 17-Mar-1999
C;Accession: A60430; A28573
R;Sletten, K.; Husebekk, A.; Husby, G.
Scand. J. Immunol. 30, 117-122, 1989
A;Title: The primary structure of equine serum amyloid a (SAA) protein.
A;Reference number: A60430; MUID:89332320; PMID:2502829
A;Accession: A60430
A;Molecule type: protein
A;Residues: 1-110 <SLE>
A;Note: 16-Leu and 44-Gln were also found
R;Sletten, K.; Husebekk, A.; Husby, G.
Scand. J. Immunol. 26, 79-84, 1987
A;Title: The amino acid sequence of an amyloid fibril protein AA isolated from the horse
A;Reference number: A28573; MUID:87291864; PMID:3616485
A;Accession: A28573
A;Molecule type: protein
A;Residues: 1-15,'L',17-77,'AB2' <SL2>
A;Note: 44-Gln was also found; the order of residues 78-79 was uncertain
C;Superfamily: amyloid protein
C;Keywords: amyloid
F;1-110/Product: serum amyloid A protein #status experimental <MAT2>
F;1-80/Product: amyloid protein A #status experimental <MAT1>

Query Match 67.3%; Score 37; DB 2; Length 110;
Best Local Similarity 83.3%; Pred. No. 9.7; Mismatches 1; Indels 0; Gaps 0;
Matches 5; Conservative 1;

QY 2 HFPHGF 7
|||||
Db 100 HFPHGF 105

RESULT 12

T28848
hypothetical protein F58F9.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
R;Nelson, J.; Wohldmann, P.
submitted to the EMBL Data Library, May 1996
A;Description: The sequence of C. elegans cosmid F58F9.
A;Reference number: Z20531
A;Accession: T28848
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-310 <NEU>
A;Cross-references: EMBL:U58756; PIDN:AAC48087.1; GSPDB:GNO0022; CESP:F58F9.1
A;Experimental source: strain Bristol N2; clone F58F9
C;Genetics:
A;Gene: CESP:F58F9.1
A;Map position: 4

Query Match 67.3%; Score 37; DB 2; Length 310;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 HFPHGF 9
|||||
Db 179 NFKPHGIS 186

RESULT 13

T19195

hypothetical protein C11E4.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T19195
R;Mortimore, B.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19087
A;Accession: T19195
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-411 <WIL>
A;Cross-references: EMBL:Z81015; PIDN:CAB02660.1; GSPDB:GNO0028; CESP:C11E4.3
A;Experimental source: clone C11E4
C;Genetics:
A;Gene: CESP:C11E4.3
A;Map position: X
A;Introns: 12/3; 91/3; 136/2; 182/3; 230/3; 307/3

Query Match 67.3%; Score 37; DB 2; Length 411;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 HFPHGF 9
|||||
Db 392 HFPHGF 399

RESULT 14

A96560
hypothetical protein F5F19.9 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: A96560
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A96560
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-642 <STO>
A;Cross-references: GB:AB005173; MID:g4220450; PIDN:AAD12677.1; GSPDB:GNO0141
C;Genetics:
A;Gene: F5F19.9
A;Map position: 1

Query Match 67.3%; Score 37; DB 2; Length 642;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHKPHGF 7
|||||
Db 157 KHKPHGF 163

RESULT 15

T29884
hypothetical protein F19G12.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T29884
R;Nhan, M.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid F19G12.
A;Reference number: Z20704

A;Accession: T29884
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-668 <NHA>
A;Cross-references: EMBL:U51997; PIDN:AAC48156.1; GSPDB:GN00028; CESP:F19G12.2
A;Experimental source: strain Bristol N2; clone F19G12
C;Genetics:
A;Gene: CESP:F19G12.2
A;Map position: X
A;Introns: 444/3; 561/2

Query Match 67.3%; Score 37; DB 2; Length 668;
Best Local Similarity 75.0%; Pred. No. 61;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HFKPHGFS 9
:|||||
Db 314 NFKPHGIS 321

Search completed: December 16, 2003, 14:19:19
Job time : 11.3333 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:09:45 ; Search time 25 Seconds
(without alignments)
92.899 Million cell updates/sec

Title: US-09-870-089B-13

Perfect score: 55

Sequence: 1 KHFKPHGFS 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	100.0	347	11 Q9ES19	Q9es19 rattus norv
2	55	100.0	351	4 Q96AQ3	Q96aq3 homo sapien
3	49	89.1	350	11 Q8CF69	Q8cf69 mus musculu
4	49	89.1	381	11 Q61328	Q61328 mus musculu
5	43	78.2	348	16 Q25732	Q25732 heliocobacte
6	43	78.2	350	16 Q9ZKA9	Q9zka9 heliocobacte
7	41	74.5	191	16 Q67912	Q67912 aquifex aeo
8	41	74.5	321	5 Q9BJW3	Q9bjw3 plasmodium
9	41	74.5	2375	5 Q8IEN3	Q8ien3 plasmodium
10	41	74.5	2391	5 Q27732	Q27732 plasmodium
11	39	70.9	375	16 Q8ZHV7	Q8zhv7 yersinia pe
12	38	69.1	51	10 Q8SMF3	Q8smf3 gossypium s
13	38	69.1	289	17 Q8Q016	Q8q016 methanosarc
14	38	69.1	317	16 Q92P43	Q92p43 rhizobium m
15	38	69.1	402	10 Q8VZW9	Q8vzw9 vigna unqui
16	38	69.1	1369	5 Q9VFW4	Q9vpw4 drosophila

17	38	69.1	1418	5 Q9GSZ6	Q9gsz6 plasmodium
18	38	69.1	1449	5 Q9GSZ7	Q9gsz7 plasmodium
19	37	67.3	128	6 Q9N0Y1	Q9n0y1 equus cabal
20	37	67.3	156	16 Q8P5G5	Q8p5g5 xanthomonas
21	37	67.3	197	11 Q9D7Q0	Q9d7q0 mus musculu
22	37	67.3	224	16 Q9CL80	Q9cl80 pasteurella
23	37	67.3	310	5 Q20993	Q20993 caenorhabdi
24	37	67.3	317	5 Q9U613	Q9u613 artemia san
25	37	67.3	411	5 Q93205	Q93205 caenorhabdi
26	37	67.3	642	10 Q23831	Q23831 arabidopsis
27	37	67.3	643	10 Q9SSV0	Q9ssv0 arabidopsis
28	37	67.3	646	10 Q9SXH0	Q9sxh0 arabidopsis
29	37	67.3	654	10 Q9MSW9	Q9msw9 arabidopsis
30	37	67.3	668	5 Q19599	Q19599 caenorhabdi
31	37	67.3	754	17 Q97U14	Q97u14 sulfolobus
32	36	65.5	248	16 Q8E8N8	Q8e8n8 shewanella
33	36	65.5	304	16 Q8XZK7	Q8xzk7 ralstonia s
34	36	65.5	362	5 Q9N882	Q9n882 plasmodium
35	36	65.5	374	5 Q9VC48	Q9vc48 drosophila
36	36	65.5	381	11 Q63095	Q63095 rattus norv
37	36	65.5	402	16 Q98MP8	Q98mp8 rhizobium l
38	36	65.5	417	2 P72417	P72417 salmonella
39	36	65.5	536	10 Q9FE45	Q9fe45 dendrobium
40	36	65.5	1229	12 Q9JIF1	Q9jif1 lymantria d
41	36	65.5	1964	10 Q9JIM2	Q9jim2 arabidopsis
42	35	63.6	38	2 Q973E7	Q973e7 helicobacte
43	35	63.6	106	13 Q98TP6	Q98tp6 platichthys
44	35	63.6	116	16 Q9K9N9	Q9k9n9 bacillus ha
45	35	63.6	120	16 Q34743	Q34743 bacillus su

ALIGNMENTS

RESULT 1

Q9ES19 PRELIMINARY; PRT; 347 AA.

AC Q9ES19, Q9ES19, TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Activating transcription factor ATF-4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20519611; PubMed=10924501;
RA Nehring R.B.; Horikawa H.P.M.; El Far O.; Kneussel M.,
RA Brandstatter J.H.; Stamm S.; Wismeyer E.; Betz H.; Karschin A.;
RT "The Metabotropic GABAB Receptor Directly Interacts with the
RT Activating Transcription Factor 4.";
RL J. Biol. Chem. 275:35185-35191(2000).
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL; AP252627; AAC31732.1; -
DR InterPro; IPR004827; TF_BZIP.
DR Pfam; PF00170; BZIP; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 347 AA; 38151 MW; CCE6BD02F263296B CRC64;

Query Match 100.0%; Score 55; DB 11; Length 347;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9

|||||

42 KHFKPHGFS 50

RESULT 2

```

Q96AQ3 PRELIMINARY; PRT; 351 AA.
ID Q96AQ3
AC Q96AQ3
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK003001; BAC25014.1; -.
SQ SEQUENCE 350 AA; 38715 MW; 6394500297E6C3CE CRC64;

Query Match 89.1%; Score 49; DB 11; Length 350;
Best Local Similarity 88.9%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
DB 42 KHLKPHGFS 50

RESULT 3
Q8CF69 PRELIMINARY; PRT; 350 AA.
ID Q8CF69
AC Q8CF69
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Activating transcription factor 4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK003001; BAC25014.1; -.
SQ SEQUENCE 350 AA; 38715 MW; 6394500297E6C3CE CRC64;

Query Match 89.1%; Score 49; DB 11; Length 350;
Best Local Similarity 88.9%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
DB 42 KHLKPHGFS 50

RESULT 4
Q61328 PRELIMINARY; PRT; 381 AA.
ID Q61328
AC Q61328
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Activating transcription factor 4 (Activating TRANSCRIPTION factor 4)
DE (ATF4).
GN ATF4 OR ATF-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=92066493; PubMed=1956797;
RA Mielnicki L.M., Pruitt S.C.;
RT "Isolation and nucleotide sequence of a murine cDNA homologous to
RT human activating transcription factor 4."
RL Nucleic Acids Res. 19:6332-6332 (1991).
RN [2]
RN SEQUENCE OF 119-381 FROM N.A.
RA Nishizawa M., Nagata S.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 119-381 FROM N.A.
RX MEDLINE=92183900; PubMed=1371974;
RA Nishizawa M., Nagata S.;
RT "cDNA clones encoding leucine-zipper proteins which interact with G-
RT CSF gene promoter element 1-binding protein."
RL FEBS Lett. 299:36-38 (1992).
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL; X61507; CAA43723.1; -.
DR TRANSFAC; T02052; -.
DR MGD; MGI:88096; Atf4.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bZIP; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 381 AA; 41810 MW; 9FC3A0CA16BB7431 CRC64;

Query Match 89.1%; Score 49; DB 11; Length 381;
Best Local Similarity 88.9%; Pred. No. 0.5;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
DB 74 KHLKPHGFS 82

RESULT 5
Q25732 PRELIMINARY; PRT; 348 AA.
ID Q25732
AC Q25732
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Cinnamyl-alcohol dehydrogenase ELI3-2 (CAD).
GN HPI104.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=2110;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Toml J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty S.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,

```

```

RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
CC -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL; AE000617; AAD08150.1; -.
DR TIGR; HP1104; -.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Hypothetical protein; Oxidoreductase; Zinc; Complete proteome.
SQ SEQUENCE 348 AA; 38645 MW; 07D7E1A63048EF48 CRC64;

Query Match 78.2%; Score 43; DB 16; Length 348;
Best Local Similarity 87.5%; Pred. No. 5.9; Mismatches 1; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 2 HFKPHGFS 9
DB 16 HFKPHDFS 23

RESULT 6
Q9ZKA9 PRELIMINARY; PRT; 350 AA.
AC Q9ZKA9;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Zinc-dependent alcohol dehydrogenase.
GN JHP1030.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL; AE001531; AAD06610.1; -.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Complete proteome.
SQ SEQUENCE 350 AA; 38939 MW; 0CC26211002E6CD1 CRC64;

Query Match 78.2%; Score 43; DB 16; Length 350;
Best Local Similarity 87.5%; Pred. No. 5.9; Mismatches 1; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 2 HFKPHGFS 9
DB 16 HFKPHDFS 23

RESULT 7
O67912 PRELIMINARY; PRT; 191 AA.
ID O67912

```

```

AC O67912;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein AQ_2159.
GN AQ_2159.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).
DR EMBL; AE000776; AAC07886.1; -.
DR InterPro; IPR001310; HIT.
DR Pfam; PF01230; HIT; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 191 AA; 22403 MW; EFA14A8699BB9165 CRC64;

Query Match 74.5%; Score 41; DB 16; Length 191;
Best Local Similarity 66.7%; Pred. No. 7.6; Mismatches 2; Indels 0; Gaps 0;
Matches 6; Conservative 2;

QY 1 KHFKPHGFS 9
DB 119 KHFKPHGFN 127

RESULT 8
Q9BJW3 PRELIMINARY; PRT; 321 AA.
ID Q9BJW3;
AC Q9BJW3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Carbamoyl phosphate synthetase II (EC 1.4.3.6) (Copper amine oxidase)
DE (Fragment).
GN CPSII.
OS Plasmodium vivax.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5855;
RN [1]
RP SEQUENCE FROM N.A.
RA Katrib M.T., Stewart T.S.;
RT "Carbamoyl phosphate synthetase II gene from Plasmodium vivax.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2).
CC -!- COFACTOR: BINDS 1 COPPER ION AND 1 TOPAQUNONE PER SUBUNIT (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE COPPER/TOPAQUNONE OXIDASE FAMILY.
DR EMBL; AF327646; AAK15525.1; -.
DR HSP; P00968; 1A9X.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR000269; CuNH_oxidase.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF02786; CPase_L_D2; 1.
DR Pfam; PF02787; CPase_L_D3; 1.
DR PRINTS; PRO00098; CPASE.
DR PROSITE; PS01164; COPPER AMINE OXID_1; 1.
DR PROSITE; PS00867; CPASE_2; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW Copper; Oxidoreductase; TPQ.
FT NON_TER 1
FT NON_TER 321

```

```

SQ SEQUENCE 321 AA; 36844 MW; 6C5001566CDC5E8A CRC64;
Query Match 74.5%; Score 41; DB 5; Length 321;
Best Local Similarity 77.8%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
|:|:|:|:|
Db 267 KYFKKHGFS 275

RESULT 9
Q81EN3 PRELIMINARY; PRT; 2375 AA.
AC Q81EN3;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Carbamoyl phosphate synthetase, putative (EC 6.3.5.5).
GN PF13 0044.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL844509; CDS2216.1; -.
RW Ligase.
KW SEQUENCE 2375 AA; 273435 MW; 398D3D59DBA24D1 CRC64;

Query Match 74.5%; Score 41; DB 5; Length 2375;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
|:|:|:|:|
Db 1154 KYFKKHGFS 1162

RESULT 10
Q27732 PRELIMINARY; PRT; 2391 AA.
AC Q27732;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Carbamoyl phosphate synthetase II (EC 6.3.5.5) (Copper amine
DE oxidase).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Flores M.V.C.;
RA STRAIN=FCQ27;
RT "Molecular Biology of Pyrimidine Biosynthesis in the Human Malarial
RT Parasite Plasmodium falciparum";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Flores M.V.C., O'Sullivan W.J., Stewart T.S.;
RA STRAIN=FCQ27;
RT "Characterisation of the Carbamoyl Phosphate Synthetase Gene from
RT Plasmodium falciparum";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2).
CC -!- COFACTOR: BINDS 1 COPPER ION AND 1 TOPAQUNONE PER SUBUNIT (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE COPPER/TOPAQUNONE OXIDASE FAMILY.

SQ SEQUENCE 2391 AA; 275674 MW; 426F3C3E612FE6E2 CRC64;

Query Match 74.5%; Score 41; DB 5; Length 2391;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KHFKPHGFS 9
|:|:|:|:|
Db 1168 KYFKKHGFS 1176

RESULT 11
Q8ZHV7 PRELIMINARY; PRT; 375 AA.
AC Q8ZHV7;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein YPO0774.
GN YPO0774 OR Y3419.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,

```

RA Petherston J.D., Lindler L.B., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of *Yersinia pestis* KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 DR EMBL; AJ414144; CAC89623.1; -;
 DR EMBL; AF013944; AM86968.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 375 AA; 41623 MW; AA1F781311AC9AB0 CRC64;

Query Match 70.9%; Score 39; DB 16; Length 375;
 Best Local Similarity 75.0%; Pred. No. 35;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 HFKPHGFS 9
 ||:||||
 Db 173 HFSYVGF 180

RESULT 12
 Q9SMF3 PRELIMINARY; PRT; 51 AA.
 AC Q9SMF3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE Microsomal omega6 desaturase enzyme (Fragment).
 GN PAD2.
 OS Gossypium stocksii.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Malvaceae; Malvaceae; Malvoideae; Gossypium.
 OX NCBI_TaxID=47602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Brubaker C.L., Liu Q., Green A.G., Marshall D.R., Singh S.P.;
 RT "Microsomal omega6 desaturase intron topologies contribute to our
 RT understanding of reticulate evolution in Gossypium (Malvaceae) and the
 RT evolution of the Australian Gossypium species.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ244917; CAB59280.1; -;
 FT NON_TER 1
 FT CHATN 1
 FT NON_TER 51
 FT NON_TER 51
 SQ SEQUENCE 51 AA; 5728 MW; B3103EFF7594E8A3 CRC64;

Query Match 69.1%; Score 38; DB 10; Length 51;
 Best Local Similarity 75.0%; Pred. No. 7.3;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHFKPHGF 8
 ||:||||
 Db 39 KPRPHGF 46

RESULT 13
 Q8Q0I6 PRELIMINARY; PRT; 289 AA.
 AC Q8Q0I6;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Putative nucleoside-diphosphate-sugar epimerase.
 GN MM0150.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strain=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
 RC MEDLINE=22120827; PubMed=12125924;
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,

RA Martinez-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C.,
 RA Bruggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 RA Fritz H.-J., Gottschalk G.;
 RT "The genome of *Methanosarcina mazei*: evidence for lateral gene
 RT transfer between Bacteria and Archaea.";
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 DR EMBL; AF013236; AM29846.1; -;
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 289 AA; 32222 MW; AF59E9997E30ADF6 CRC64;

Query Match 69.1%; Score 38; DB 17; Length 289;
 Best Local Similarity 66.7%; Pred. No. 41;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KHFKPHGFS 9
 ||:||||
 Db 249 KHYDEHGF 257

RESULT 14
 Q92P43 PRELIMINARY; PRT; 317 AA.
 AC Q92P43;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein R01950.
 GN R01950 OR SMC04290.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591789; CAC46529.1; -;
 DR InterPro; IPR004304; FmdA Amda.
 DR Pfam; PF03069; FmdA_Amda_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 317 AA; 33877 MW; 7705D716F6B9BE2A CRC64;

Query Match 69.1%; Score 38; DB 16; Length 317;
 Best Local Similarity 75.0%; Pred. No. 45;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHFKPHGF 8
 ||:||||
 Db 89 EHFKPSGF 96

RESULT 15
 Q8VZW9 PRELIMINARY; PRT; 402 AA.
 AC Q8VZW9;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Succinoaminoimidazolecarboxamide ribonucleotide synthetase
 DE (EC 6.3.2.6).
 GN PUR7.
 OS Vigna unguiculata (Cowpea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX NCBI_TaxID=3917;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Vita 3;
RA Smith P.M.C., Mann A.J., Reck V.;
RT "vupur7 mRNA from cowpea encoding succinoaminoimidazolecarboximide
ribonucleotide synthetase.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + 5-AMINO-1-(5-PHOSPHO-D-
RIBOSYL)IMIDAZOLE-4-CARBOXYLATE + L-ASPARTATE = ADP + PHOSPHATE +
(S)-2-[5-AMINO-1-(5-PHOSPHO-D-RIBOSYL)IMIDAZOLE-4-
CARBOXAMIDO]SUCCINATE.
CC -!- PATHWAY: DE NOVO PURINE BIOSYNTHESIS; SEVENTH STEP.
CC -!- SIMILARITY: BELONGS TO THE SAICAR SYNTHETASE FAMILY.
DR EMBL; AY063233; AAL48317.1; -.
DR InterPro; IPR001636; SAICAR_synt.
DR Pfam; PF01259; SAICAR_synt; 1.
DR PRODOM; PD003043; SAICAR_synt; 1.
DR PROSITE; PS01057; SAICAR_SYNTHETASE_1; 1.
DR PROSITE; PS01058; SAICAR_SYNTHETASE_2; 1.
KW Ligase; Purine biosynthesis.
SQ SEQUENCE 402 AA; 45054 MW; 23FCB2DC29BAE3BD CRC64;

Query Match 69.1%; Score 38; DB 10; Length 402;
Best Local Similarity 100.0%; Pred.No. 57;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FKPHGF 8
| | | | |
Db 38 FKPHGF 43

Search completed: December 16, 2003, 14:18:04
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:05 ; Search time 33 Seconds
(without alignments)
43.289 Million cell updates/sec

Title: US-09-870-089B-3
Perfect score: 57
Sequence: 1 FLYKWHGFV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	100.0	9	23 AAU74681	Human cancer antig
2	43	75.4	690	22 ABB61794	Drosophila melanog
3	41	71.9	9	23 AAU74684	Human cancer antig
4	40	70.2	683	22 ABB59504	Drosophila melanog
5	40	70.2	693	17 AAR89114	Bombyx mori (pro)p
6	38	66.7	184	22 AAU19356	Human G protein-co
7	38	66.7	220	22 AAU25704	G protein-coupled
8	38	66.7	448	23 ABB92481	Herbicidally activ
9	37	64.9	9	23 AAU74683	Human cancer antig

10	37	64.9	23	20	AAU11930	Human 5' EST secre
11	37	64.9	92	22	AAG74446	Human colon cancer
12	37	64.9	22	22	ABB11269	Human LCP homolog
13	37	64.9	151	22	ABB66825	Drosophila melanog
14	37	64.9	517	22	ABB69124	Drosophila melanog
15	37	64.9	524	22	ABB91229	Listeria monocytog
16	37	64.9	552	23	ABB49129	Drosophila melanog
17	37	64.9	684	22	ABB57998	S. epidermidis ope
18	37	64.9	754	22	AAG82101	Human angiotensin
19	36	63.2	14	22	AA000213	Human secreted pro
20	36	63.2	51	22	AAB64578	Human secreted pro
21	36	63.2	52	21	AAG42515	Arabidopsis thalia
22	36	63.2	52	21	AAG44592	Arabidopsis thalia
23	36	63.2	52	21	AAG61759	Arabidopsis thalia
24	36	63.2	87	22	AA002856	Human polypeptide
25	36	63.2	144	20	AAW73761	M. tuberculosis an
26	36	63.2	144	20	AAW73651	M. tuberculosis an
27	36	63.2	144	22	AAU08121	Mycobacterium tube
28	36	63.2	144	23	ABG77968	M. tuberculosis CD
29	36	63.2	197	20	AAU29080	T. gondii immunoge
30	36	63.2	197	22	AAU25551	T. gondii immunoge
31	36	63.2	275	22	AAU73667	Mouse ageing inhib
32	36	63.2	295	22	AAU73666	Mouse ageing inhib
33	36	63.2	348	18	AAU27267	Streptococcus pneu
34	36	63.2	348	20	AAU29981	Streptococcus pneu
35	36	63.2	400	22	AAU95039	Human protein sequ
36	36	63.2	424	22	ABG06198	Novel human diagno
37	36	63.2	448	23	ABB91307	Herbicidally activ
38	36	63.2	480	20	AAU09552	Streptococcus pneu
39	36	63.2	480	21	AAU32829	Glutamyl tRNA synt
40	36	63.2	486	22	AAU37862	Streptococcus pneu
41	36	63.2	486	24	AAU02579	S. pneumoniae type
42	36	63.2	685	18	AAU14441	Prophenol oxidase
43	36	63.2	695	23	ABP39496	Staphylococcus epi
44	36	63.2	1014	22	AAU73665	Mouse ageing inhib
45	35	61.4	1043	22	AAU73664	Mouse ageing inhib
			77	22	AAU21269	Human novel foetal

ALIGNMENTS

RESULT 1

AAU74681

ID AAU74681 standard; Peptide; 9 AA.

XX AAU74681;

XX AC

XX DT 09-APR-2002 (first entry)

XX DE Human cancer antigen ATF4/CREB-2 based immunogenic ligand #1.

XX DE Human cancer antigen ATF4/CREB-2 based immunogenic ligand #1.

XX DE Human; cancer antigen; ATF4; CREB-2; vaccine; cytostatic;

XX DE immunogenic ligand; gene therapy; MHC; major histocompatibility complex;

XX DE adoptive immunotherapy; cancer; ovarian cancer.

XX OS Homo sapiens.

XX OS Synthetic.

XX XX WO200192306-A2.

XX PN

XX PD 06-DEC-2001.

XX XX

XX PF 30-MAY-2001; 2001WO-US17454.

XX PR

XX PR 31-MAY-2000; 2000US-209388P.

XX PR 20-DEC-2000; 2000US-257007P.

XX PA (GENZ) GENZYME CORP.

XX XX

XX PI Nicolette CA;

XX XX

XX DR WPI; 2002-097764/13.

XX XX

CC Also included are an antibody that specifically recognises and binds the
 CC compound, a method for inducing an immune response in a subject by
 CC delivering the compound, a method of immunotherapy comprising
 CC administering to a subject the antibody, an immune effector cell that has
 CC been raised in vitro or in vivo in the presence and at the expense of an
 CC antigen presenting cell that presents the immunogenic compound in the
 CC context of an MHC (major histocompatibility complex) molecule and a
 CC method of adoptive immunotherapy comprising administering the immune
 CC effector cell. The compounds are useful for modulating an immune response
 CC to the synthetic and naturally occurring compounds. The compounds are
 CC especially useful in gene therapy or as components of anti-cancer
 CC vaccines. The compounds are useful for treating cancer, particularly
 CC ovarian cancer. The compounds are also useful for generating antibodies
 CC that specifically recognise and bind to these molecules. These
 CC antibodies are further useful for immunotherapy when administered to a
 CC subject. The peptides, polypeptides and polynucleotides are useful in
 CC diagnostic methods, for the detection and purification of antibodies,
 CC or as immunogens for the production of antibodies. The present
 CC sequence represents a human cancer antigen ATF4/CREB-2 based immunogenic
 CC ligand of the invention.
 CC Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
 CC encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
 CC but have not been cross-referenced or CDS features put in due to the
 CC degeneracy of the DNA sequences.

XX Sequence 9 AA;

Query Match 71.9%; Score 41; DB 23; Length 9;
 Best Local Similarity 66.7%; Pred. NO. 9.3e+05;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLYKWHGFV 9

DB 1 FLKWHWYV 9

RESULT 4

ABBS9504
 ID ABB59504 standard; Protein; 683 AA.

XX AC ABB59504;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 5304.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL03607.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX PS Disclosure; SEQ ID NO 5304; 21bp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 683 AA;

Query Match 70.2%; Score 40; DB 22; Length 683;
 Best Local Similarity 71.4%; Pred. NO. 1.2e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 YKWHGFV 9

DB 403 YKWHAFI 409

RESULT 5

AAR89114
 ID AAR89114 standard; Protein; 693 AA.

XX AC AAR89114;

XX DT 19-JUN-1996 (first entry)

XX DE Bombyx mori (pro)phenoloxidase Phe52-Val693.

XX KW Silkworm; larvae; pro; phenoloxidase; Phe52-Val693; Phe52-Gly685
 KW labelling oxidase.

XX OS Bombyx mori.

XX FH Key Location/Qualifiers

FT Peptide 25..51

FT Peptide /note= "peptide fragment"

FT Peptide 52..61

FT Peptide /note= "peptide fragment"

XX PN JP07289251-A.

XX PD 07-NOV-1995.

XX PF 22-APR-1994; 94JP-0085096.

XX PR 22-APR-1994; 94JP-0085096.

XX PA (WAKP) WAKO PURE CHEM IND LTD.

XX DR WPI; 1996-015259/02.

XX DR N-PSDB; AAT10240.

XX PT (Pro)phenol:oxidase from silkworm - useful as a labelling oxidase

XX PS Claim 1; Pages 11-14; 29pp; Japanese.

CC The B. mori silkworm larvae derived (pro)phenoloxidases Phe52-Val693
 CC (AAR89114) and Phe52-Gly685 (AAR89115), are encoded by AAT10240 and
 CC AAT10241, respectively. The (pro)phenoloxidases can be used as new
 CC labelling oxidases.

XX Sequence 693 AA;

Query Match 70.2%; Score 40; DB 17; Length 693;
 Best Local Similarity 44.4%; Pred. NO. 1.3e+02;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGFV 9

CC atherosclerosis, cancer, and sexual dysfunction. AAU25617-AAU25726
 CC represent the amino acid sequences of novel human G protein-coupled
 CC receptors, nGPCR-2031 to nGPCR-2140 respectively, as described in the
 CC invention.

XX SQ Sequence 220 AA;
 Query Match 66.7%; Score 38; DB 22; Length 220;
 Best Local Similarity 66.7%; Pred. No. 87;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 FLYKWHGFV 9
 Db 84 FLNLWHGFL 92
 ||| |||:

RESULT 8
 ABB92481
 ID ABB92481 standard; Protein; 448 AA.

XX AC ABB92481;
 XX DT 31-MAY-2002 (first entry)
 XX DE Herbicidally active polypeptide SEQ ID NO 1692.
 XX KW Herbicidal; plant; agriculture; herbicide.
 XX OS Arabidopsis thaliana.
 XX PN WO200210210-A2.
 XX PD 07-FEB-2002.
 XX PF 28-AUG-2001; 2001WO-EP09892.
 XX PR 28-AUG-2001; 2001WO-EP09892.
 XX PA (FARB) BAYER AG.
 XX PI Tietjen K, Weidler M;
 XX DR WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -

PS Claim 5; SEQ ID NO 1692; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX SQ Sequence 448 AA;
 Query Match 66.7%; Score 38; DB 23; Length 448;
 Best Local Similarity 75.0%; Pred. No. 1.8e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLYKWHGF 8
 Db 8 FLYPWFGF 15
 ||| |||

RESULT 9

AAU74683
 ID AAU74683 standard; Peptide; 9 AA.
 XX AC AAU74683;
 XX DT 09-APR-2002 (first entry)

XX DE Human cancer antigen ATF4/CREB-2 based immunogenic ligand #3.
 XX KW Human; cancer antigen; ATF4; CREB-2; vaccine; cytostatic;
 KW immunogenic ligand; gene therapy; MHC; major histocompatibility complex;
 KW adoptive immunotherapy; cancer; ovarian cancer.
 XX OS Homo sapiens.
 OS Synthetic.
 XX PN WO200192306-A2.
 XX PD 06-DEC-2001.
 XX PF 30-MAY-2001; 2001WO-US17454.
 XX PR 31-MAY-2000; 2000US-209388P.
 XX PR 20-DEC-2000; 2000US-257007P.
 XX PA (GENZ) GENZYME CORP.

XX PI Nicolette CA;
 XX DR WPI; 2002-097764/13.

XX New therapeutic compounds comprising immunogenic ligands, useful for
 PT modulating an immune response, particularly for treating ovarian
 PT cancer, and as components of anticancer vaccines -

PS Claim 3; Page 55; 68pp; English.

XX The invention relates to compounds comprising an immunogenic ligand
 CC whose sequence is based in part on residues 42-50 of human cancer
 CC antigen ATF4/CREB-2 (not defined) and the polynucleotides encoding them.
 CC Also included are an antibody that specifically recognises and binds the
 CC compound, a method for inducing an immune response in a subject by
 CC delivering the compound, a method of immunotherapy comprising
 CC administering to a subject the antibody, an immune effector cell that has
 CC been raised in vitro or in vivo in the presence and at the expense of an
 CC antigen presenting cell that presents the immunogenic compound in the
 CC context of an MHC (major histocompatibility complex) molecule and a
 CC method of adoptive immunotherapy comprising administering the immune
 CC effector cell. The compounds are useful for modulating an immune response
 CC to the synthetic and naturally occurring compounds. The compounds are
 CC especially useful in gene therapy or as components of anti-cancer
 CC vaccines. The compounds are useful for treating cancer, particularly
 CC ovarian cancer. The compounds are also useful for generating antibodies
 CC that specifically recognise and bind to these molecules. These
 CC antibodies are further useful for immunotherapy when administered to a
 CC subject. The peptides, polypeptides and polynucleotides are useful in
 CC diagnostic methods, for the detection and purification of antibodies,
 CC or as immunogens for the production of antibodies. The present
 CC sequence represents a human cancer antigen ATF4/CREB-2 based immunogenic
 CC ligand of the invention.

CC Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
 CC encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
 CC but have not been cross-referenced or CDS features put in due to the
 CC degeneracy of the DNA sequences.

XX SQ Sequence 9 AA;

Query Match 64.9%; Score 37; DB 23; Length 9;
 Best Local Similarity 66.7%; Pred. No. 9.3e+05;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLYKWHGFV 9
 ||| |||

XX Human LCP homologue, SEQ ID NO:1639.

DE Human; cytokine; cell proliferation; cell differentiation; growth factor;

XX haematopoiesis regulation; tissue growth; immunomodulator; activin;

KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;

KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;

KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;

KW chronic inflammatory condition; proliferative retinopathy;

KW atherosclerosis; coronary heart disease; arterial ischaemia;

KW bone disorder; osteoporosis; vascular growth disorder;

KW tissue regeneration; wound healing; infection; immune disorder;

KW cell culture; drug screening; gene therapy; antiinflammatory;

KW cytoskeletal; antiarthritic; haemostatic; antiarteriosclerotic;

KW cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;

KW antifungal; vulnery; antiulcer.

XX Homo sapiens.

OS WO200157188-A2.

XX 09-AUG-2001.

PD 05-FEB-2001; 2001WO-US03800.

PF 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

DR N-PSDB; ABA08513.

XX Human proteins and DNA encoding sequences useful for preventing,

PT treating or ameliorating a medical condition in a mammalian subject

PT e.g. arthritis and cancer -

XX Claim 20; Page 161; 1963pp; English.

PS Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and

CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The

CC invention also relates to vectors and recombinant host cells comprising a

CC nucleotide of the invention, methods of producing the novel polypeptides,

CC antibodies against the polypeptides, methods of detecting the nucleotides,

CC or polypeptides in a sample, and methods of identifying compounds which

CC bind to polypeptides of the invention. Although novel, many of the

CC polypeptides of the invention have homology to known proteins, thereby

CC giving an insight into their probable biological activities, and hence

CC potential therapeutic applications. The polypeptides of the invention may

CC have various activities, including cytokine, cell proliferation or cell

CC differentiation activities; stem cell growth factor activity;

CC haematopoiesis regulatory activity; tissue growth activity;

CC immunomodulatory activity; activin- or inhibin-related activities;

CC chemotactic or chemokinetic activities; haemostatic, thrombotic or

CC thrombolytic activities; receptor or ligand activation; or may be

CC involved in oncogenesis, cancer cell proliferation or metastasis.

CC Depending on their biological activities, polypeptides and nucleotides of

CC the invention are useful for preventing, treating or ameliorating medical

CC conditions, e.g., by protein or gene therapy. Such conditions include

CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell

CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),

CC proliferative retinopathy, atherosclerosis, coronary heart disease,

CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal

CC vascular growth. Polypeptides involved with tissue regeneration and

CC repair (or nucleic acids encoding them) may be used to promote wound

CC healing (e.g., of burns, incisions and ulcers), while those with

CC immunomodulatory activities may be used in the treatment of viral,

CC bacterial and fungal infections in addition to immune disorders.

CC Polypeptides with growth factor activity may be used in cell cultures to

CC promote cell growth. For example, such polypeptides may be used to

CC manipulate stem cells in culture to give rise to neuroepithelial cells

CC that can be used to augment or replace cells damaged by illness,

CC autoimmune disease or accidental damage. The polypeptides and nucleotides

CC may also be used in the diagnosis of the above conditions, and in drug

CC screening techniques. The present sequence represents a novel human

CC polypeptide of the invention.

XX Sequence 151 AA;

SQ Query Match 64.9%; Score 37; DB 22; Length 151;

Best Local Similarity 66.7%; Pred. No. 89;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 FLYKWHGVF 9

DB ||||| |

35 FLYVWGGYV 43

RESULT 13

ABB66825

ID ABB66825 standard; Protein; 517 AA.

XX AC ABB66825;

XX 26-MAR-2002 (first entry)

DT XX Drosophila melanogaster polypeptide SEQ ID NO 27267.

DE XX Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

PD XX 23-MAR-2001; 2001WO-US09231.

PF XX 23-MAR-2000; 2000US-191637P.

PR XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

XX N-PSDB; ABL10928.

DR New isolated nucleic acid detection reagent for detecting 1000 or more

CC genes from Drosophila and for elucidating cell signalling and cell-cell

CC interactions -

XX Disclosure; SEQ ID NO 27267; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 517 AA;

SQ Query Match 64.9%; Score 37; DB 22; Length 517;

Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 YKWHG 7

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:11:00 ; Search time 12.3333 Seconds
(without alignments)
30.875 Million cell updates/sec

Title: US-09-870-089B-3
Perfect score: 57
Sequence: 1 FLYKWHGFV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/iaa/5A COMB pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1 pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	63.2	144	US-09-073-009-13	Sequence 13, Appl
2	36	63.2	197	US-09-216-393B-330	Sequence 330, App
3	36	63.2	348	US-08-844-153-2	Sequence 2, Appli
4	36	63.2	453	US-09-252-991A-31268	Sequence 31268, A
5	36	63.2	463	US-09-252-991A-28820	Sequence 28820, A
6	36	63.2	480	US-08-962-203-2	Sequence 2, Appli
7	36	63.2	480	US-09-282-125A-2	Sequence 2, Appli
8	36	63.2	480	US-09-273-142-2	Sequence 2, Appli
9	36	63.2	695	US-09-134-001C-4341	Sequence 4341, Ap
10	35	61.4	351	US-09-632-947B-11	Sequence 11, Appl
11	35	61.4	472	US-09-107-532A-6117	Sequence 6117, Ap
12	34	59.6	257	US-09-407-532A-5440	Sequence 5440, Ap
13	34	59.6	452	US-09-252-991A-25679	Sequence 25679, A
14	34	59.6	539	US-09-328-352-5771	Sequence 5771, Ap
15	34	59.6	629	US-09-252-991A-31575	Sequence 31575, A
16	34	59.6	801	US-09-252-991A-27870	Sequence 27870, A
17	33	57.9	82	US-09-482-273-185	Sequence 185, App
18	33	57.9	82	US-09-482-273-222	Sequence 222, App
19	33	57.9	103	US-09-482-273-219	Sequence 219, App
20	33	57.9	104	US-09-482-273-182	Sequence 182, App
21	33	57.9	124	US-08-591-632-89	Sequence 89, Appl
22	33	57.9	124	US-08-591-632-90	Sequence 90, Appl
23	33	57.9	124	US-08-591-632-91	Sequence 91, Appl
24	33	57.9	124	US-08-591-632-92	Sequence 92, Appl
25	33	57.9	124	US-09-611-451-89	Sequence 89, Appl
26	33	57.9	124	US-09-611-451-90	Sequence 90, Appl
27	33	57.9	124	US-09-611-451-91	Sequence 91, Appl

Sequence 92, Appl
Sequence 183, App
Sequence 220, App
Sequence 119, App
Sequence 38, Appl
Sequence 5216, Ap
Sequence 6294, Ap
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 30025, A
Sequence 19865, A
Sequence 4435, Ap
Sequence 35, Appl
Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-09-073-009-13
; Sequence 13, Application US/09073009
; Patent No. 6555653
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
; US-09-073-009-13

Query Match 63.2%; Score 36; DB 4; Length 144;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 1; Mismatches 2; Indels C; Gaps 0;
QY 1 FLYKWHGF 8

```
Db      134 FLCSWHG 141
      || |||
      TOPOLOGY: linear
      MOLECULE TYPE: protein
US-08-844-153-2
Query Match      63.2%; Score 36; DB 2; Length 348;
Best Local Similarity 62.5%; Pred. No. 77;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      2 LYKWHGFV 9
Db      159 IYKWHDMV 166
      :|||||
      RESULT 4
US-09-252-991A-31268
; Sequence 31268, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31268
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31268
Query Match      63.2%; Score 36; DB 4; Length 453;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LYKWHGFV 9
Db      312 LYKNGSV 319
      :|||||
      RESULT 5
US-09-252-991A-28820
; Sequence 28820, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28820
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28820
Query Match      63.2%; Score 36; DB 4; Length 463;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 FLYKWHGFV 9
Db      306 FRERWHGFL 314
      :|||||
```

```
Db      134 FLCSWHG 141
      || |||
      TOPOLOGY: linear
      MOLECULE TYPE: protein
US-08-844-153-2
Query Match      63.2%; Score 36; DB 4; Length 197;
Best Local Similarity 71.4%; Pred. No. 43;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LYKWHGF 8
Db      173 LLRWGHF 179
      :|||||
      RESULT 3
US-08-844-153-2
; Sequence 2, Application US/08844153
; Patent No. 5958734
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: NO. 5958734el Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/844,153
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9607992.6
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: F31459
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```


RESULT 6
US-08-962-203-2
; Sequence 2, Application US/08962203
; Patent No. 5976840
; GENERAL INFORMATION:
; APPLICANT: Jaworski, Deborah
; APPLICANT: Lawlor, Elizabeth
; APPLICANT: Wang, Min
; TITLE OF INVENTION: NOVEL STREPTOCOCCAL ERS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/962,203
; FILING DATE: 31-OCT-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/844,153
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: 9607992.6
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31459-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-962-203-2

Query Match 63.2%; Score 36; DB 2; Length 480;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LYKWHGFV 9
:|||||
Db 159 IYKWHDMV 166

RESULT 7
US-09-282-125A-2
; Sequence 2, Application US/09282125A
; Patent No. 6165760
; GENERAL INFORMATION:
; APPLICANT: Jaworski, Deborah
; APPLICANT: Lawlor, Elizabeth
; APPLICANT: Wang, Min
; TITLE OF INVENTION: NOVEL STREPTOCOCCAL ERS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia

STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/282,125A
FILING DATE: 07-Aug-2000
CLASSIFICATION: <Unknown>
18-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/962,203
FILING DATE: <Unknown>
APPLICATION NUMBER: 9607992.6
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31459-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX: <Unknown>
18-APR-1996
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-282-125A-2

Query Match 63.2%; Score 36; DB 3; Length 480;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LYKWHGFV 9
:|||||
Db 159 IYKWHDMV 166

RESULT 8
US-09-273-142-2
; Sequence 2, Application US/09273142
; Patent No. 6300119
; GENERAL INFORMATION:
; APPLICANT: Jaworski, Deborah
; APPLICANT: Lawlor, Elizabeth
; APPLICANT: Wang, Min
; TITLE OF INVENTION: NOVEL STREPTOCOCCAL ERS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/273,142
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: 08/962,203
; FILING DATE:
; PRIOR APPLICATION NUMBER: 9607992.6
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31459-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-273-142-2

```

```

Query Match      63.2%; Score 36; DB 4; Length 480;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      2 LYKWHGFV 9
Db      159 IYKWHDMV 166

```

RESULT 9

```

US-09-134-001C-4341
; Sequence 4341, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4341
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4341

```

```

Query Match      63.2%; Score 36; DB 4; Length 695;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 FLYKWHGF 8
Db      413 FVQWHGF 420

```

RESULT 10

```

US-09-632-947B-11
; Sequence 11, Application US/09632947B
; Patent No. 6356845
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: CRYSTALLIZATION AND STRUCTURE DETERMINATION OF
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: UDP-N-ACETYLENOLPYRUVYLGLUCOSAMINE REDUCTASE (S. aureus
; TITLE OF INVENTION: MurB)
; FILE REFERENCE: 268.6241 0101
; CURRENT APPLICATION NUMBER: US/09/632,947B

```

```

; CURRENT FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/147,164
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Bordetella pertussis
; US-09-632-947B-11

```

```

Query Match      61.4%; Score 35; DB 4; Length 351;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      5 WHGFV 9
Db      105 WHGFV 109

```

RESULT 11

```

US-09-107-532A-6117
; Sequence 6117, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:

```

```

; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277

```

```

; INFORMATION FOR SEQ ID NO: 6117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:

```

```

; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...472
; SEQUENCE DESCRIPTION: SEQ ID NO: 6117:
; US-09-107-532A-6117

```

```

Query Match      61.4%; Score 35; DB 4; Length 472;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;

```

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLYKWH 6
Db 183 YLYRWH 188

RESULT 12
US-09-107-532A-5440
; Sequence 5440, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5440:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...257
; SEQUENCE DESCRIPTION: SEQ ID NO: 5440:
US-09-107-532A-5440

Query Match 59.6%; Score 34; DB 4; Length 257;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLYKWH 6
Db 11 YLYRWH 16

RESULT 13
US-09-252-991A-25679
; Sequence 25679, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLYKWH 6
Db 11 YLYRWH 16

RESULT 14
US-09-328-352-5771
; Sequence 5771, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5771
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5771

Query Match 59.6%; Score 34; DB 4; Length 539;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 WHGFV 9
Db 79 WHGFI 83

RESULT 15
US-09-252-991A-31575
; Sequence 31575, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31575
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31575

Query Match 59.6%; Score 34; DB 4; Length 629;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 5 WHGFV 9
Db 172 WHGFI 176

Search completed: December 16, 2003, 14:20:39
Job time : 13.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:18:11 ; Search time 22.1667 Seconds
(without alignments)
75.512 Million cell updates/sec

Title: US-09-870-089B-3
Perfect score: 57
Sequence: 1 FLYKWHGFV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	57	100.0	9	12	US-09-870-089B-3	Sequence 3, Appli
2	41	71.9	9	12	US-09-870-089B-9	Sequence 9, Appli
3	38	66.7	220	11	US-09-791-279-198	Sequence 198, App
4	37	64.9	9	12	US-09-870-089B-7	Sequence 7, Appli
5	37	64.9	92	15	US-10-106-698-5220	Sequence 5220, Ap
6	36	63.2	144	9	US-09-073-009-13	Sequence 13, Appl
7	36	63.2	144	9	US-09-023-588-13	Sequence 13, Appl
8	36	63.2	144	9	US-09-793-306-13	Sequence 13, Appl
9	36	63.2	197	9	US-09-216-393-330	Sequence 330, App
10	36	63.2	197	12	US-10-321-856-330	Sequence 330, App
11	36	63.2	486	9	US-09-815-242-13455	Sequence 13455, A
12	35	61.4	96	9	US-09-925-299-1180	Sequence 1180, Ap
13	35	61.4	96	11	US-09-925-299-1180	Sequence 1180, Ap
14	35	61.4	99	9	US-09-815-248-12	Sequence 12, Appl
15	35	61.4	148	12	US-10-407-079-109	Sequence 109, App

16	35	61.4	265	15	US-10-203-224-5	Sequence 5, Appli
17	35	61.4	302	9	US-09-443-704-12	Sequence 12, Appl
18	35	61.4	302	14	US-10-008-118A-12	Sequence 12, Appl
19	35	61.4	332	10	US-10-166-225A-53	Sequence 53, Appl
20	35	61.4	351	12	US-09-991-211-11	Sequence 11, Appl
21	35	61.4	453	15	US-10-156-761-8288	Sequence 8288, Ap
22	35	61.4	749	15	US-10-156-761-10326	Sequence 10326, A
23	35	61.4	1421	14	US-10-108-605-323	Sequence 1256, App
24	34	59.6	77	9	US-09-867-550-1256	Sequence 11, Appl
25	34	59.6	270	9	US-09-823-356-11	Sequence 4775, Ap
26	34	59.6	281	15	US-10-106-698-4775	Sequence 4775, Ap
27	34	59.6	291	15	US-10-156-761-14794	Sequence 14794, A
28	34	59.6	303	15	US-10-156-761-7706	Sequence 7706, Ap
29	34	59.6	320	12	US-10-094-749-3096	Sequence 3096, Ap
30	34	59.6	321	10	US-09-738-626-3676	Sequence 3676, Ap
31	34	59.6	333	15	US-10-156-761-12108	Sequence 12108, A
32	34	59.6	418	12	US-10-354-804-18	Sequence 18, Appl
33	34	59.6	531	9	US-09-815-242-5114	Sequence 5114, Ap
34	34	59.6	586	12	US-10-177-478-3	Sequence 3, Appli
35	34	59.6	586	15	US-10-176-884-12	Sequence 12, Appl
36	34	59.6	596	15	US-10-213-990-51	Sequence 51, Appl
37	34	59.6	610	10	US-09-792-630-21	Sequence 21, Appl
38	34	59.6	610	11	US-09-953-351-21	Sequence 21, Appl
39	34	59.6	610	14	US-10-080-376-21	Sequence 21, Appl
40	34	59.6	610	15	US-10-082-671-27	Sequence 27, Appl
41	34	59.6	610	15	US-10-097-100-21	Sequence 21, Appl
42	34	59.6	610	15	US-10-023-208-21	Sequence 21, Appl
43	33	57.9	17	12	US-10-189-437-566	Sequence 566, App
44	33	57.9	25	12	US-10-189-437-565	Sequence 565, App
45	33	57.9	26	15	US-10-097-065-365	Sequence 365, App

ALIGNMENTS

RESULT 1

US-09-870-089B-3
; Sequence 3, Application US/09870089B
; Publication NO. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 6812681209900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATP4/CREB-2
US-09-870-089B-3

Query Match 100.0%; Score 57; DB 12; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLYKWHGFV 9
| | | | | | | | | |
Db 1 FLYKWHGFV 9

RESULT 2

US-09-870-089B-9
; Sequence 9, Application US/09870089B
; Publication NO. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 6812681209900
; CURRENT APPLICATION NUMBER: US/09/870,089B

; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATF4/CREB-2
US-09-870-089B-9

Query Match 71.9%; Score 41; DB 12; Length 9;
Best Local Similarity 66.7%; Pred. No. 6.1e+05;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLYKWHGFV 9
||:||||:
Db 1 FLHKWHWV 9

RESULT 3
US-09-791-279-198
; Sequence 198, Application US/09791279
; Publication No. US20030050456A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; APPLICANT: Wood, Linda S.
; APPLICANT: Parodi, Luis
; APPLICANT: Lind, Peter
; TITLE OF INVENTION: No. US20030050456A1e1 G Protein-Coupled Receptors
; FILE REFERENCE: 00048.US1
; CURRENT APPLICATION NUMBER: US/09/791,279
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/184,715
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184725
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,712
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,606
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,602
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,604
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,822
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,710
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,689
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,690
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,716
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 198
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-279-198

Query Match 66.7%; Score 38; DB 11; Length 220;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLYKWHGFV 9
||:||||:
Db 84 FLNLWHGFL 92

RESULT 4

US-09-870-089B-7
; Sequence 7, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 6812688120900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATF4/CREB-2
US-09-870-089B-7

Query Match 64.9%; Score 37; DB 12; Length 9;
Best Local Similarity 66.7%; Pred. No. 6.1e+05;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLYKWHGFV 9
||:||||:
Db 1 FLHKWHWV 9

RESULT 5
US-10-106-698-5220
; Sequence 5220, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5220
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5220

Query Match 64.9%; Score 37; DB 15; Length 92;
Best Local Similarity 71.4%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LYKWHGF 8
||:||||:
Db 6 LFKWHNF 12

RESULT 6
US-09-073-009-13
; Sequence 13, Application US/09073009
; Patent No. US20010012888A1
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, David C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:

ADDRESSES: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
US-09-073-009-13

Query Match 63.2%; Score 36; DB 9; Length 144;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGF 8
DB 134 FLCSWHGY 141

RESULT 7
US-09-023-588-13
; Sequence 13, Application US/09023588
; Patent No. US20020081579A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark R.
; TITLE OF INVENTION: METHOD FOR THE ISOLATION OF NOVEL ANTIGENS
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,588
; FILING DATE: 14-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.445
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
US-09-023-588-13

Query Match 63.2%; Score 36; DB 9; Length 144;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGF 8
DB 134 FLCSWHGY 141

RESULT 8
US-09-793-306-13
; Sequence 13, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Ovendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Tb224
; NAME/KEY: MOD RES
; LOCATION: (1)-(144)
; OTHER INFORMATION: Xaa = any amino acid
US-09-793-306-13

Query Match 63.2%; Score 36; DB 9; Length 144;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGF 8
DB 134 FLCSWHGY 141

RESULT 9
US-09-216-393-330
; Sequence 330, Application US/09216393
; Patent No. US2001001447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: TX-1-C2

```
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 330
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-09-216-393-330

Query Match      63.2%; Score 36; DB 9; Length 197;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LYKWHGF 8
      | :| | | |
Db      173 LLRWGHF 179

RESULT 10
US-10-321-856-330
; Sequence 330, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THEREOF
; FILE REFERENCE: TX-1-C2-1
; CURRENT APPLICATION NUMBER: US/10/321,856
; CURRENT FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 09/216,393
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/994,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 330
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-10-321-856-330

Query Match      63.2%; Score 36; DB 12; Length 197;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LYKWHGF 8
      | :| | | |
Db      173 LLRWGHF 179

RESULT 11
US-09-815-242-13455
; Sequence 13455, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Toxoplasma gondii
; FILE REFERENCE: ELTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

```
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13455
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13455

Query Match      63.2%; Score 36; DB 9; Length 486;
Best Local Similarity 62.5%; Pred. No. 4.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      2 LYKWHGFV 9
      | :| | | |
Db      165 IYKWHDMV 172

RESULT 12
US-09-925-299-1180
; Sequence 1180, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1180
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (42)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (50)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (81)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (95)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1180

Query Match      61.4%; Score 35; DB 9; Length 96;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```


QY 3 YKWHGF 8
Db 69 YKQGF 74

RESULT 13

US-09-925-299-1180
; Sequence 1180, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1180
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (42)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (50)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (81)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (95)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

Query Match 61.4%; Score 35; DB 11; Length 96;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 YKWHGF 8
Db 69 YKQGF 74

RESULT 14

US-09-815-248-12
; Sequence 12, Application US/09815248
; Patent No. US20020098540A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: PENNICA, DIANE
; TITLE OF INVENTION: NOVEL POLYPEPTIDES, AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: 10716.36
; CURRENT APPLICATION NUMBER: US/09/815,248
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/191,258
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 99
; TYPE: PRT

; ORGANISM: Drosophila sp.
US-09-815-248-12

Query Match 61.4%; Score 35; DB 9; Length 99;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LYKWHGFV 9
Db 13 LYIWHKFV 20

RESULT 15

US-10-407-079-109
; Sequence 109, Application US/10407079
; Publication No. US20030215860A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Glucksman, Maria A.
; APPLICANT: Silos-Santiago, Inmaculada
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Galvin, Katherine M.
; TITLE OF INVENTION: 18636, 2466, 43238, 1983, 52881, 2398,
; FILE REFERENCE: MPI03-0510NMIM
; CURRENT APPLICATION NUMBER: US/10/407,079
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US 10/226,102
; PRIOR FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US 60/314,041
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US 10/225,094
; PRIOR FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: US 60/314,185
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US 10/272,417
; PRIOR FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/715,790
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: US 60/191,845
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 10/282,837
; PRIOR FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: US 09/796,338
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,059
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 109
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Amino acid consensus sequence

US-10-407-079-109

Query Match 61.4%; Score 35; DB 12; Length 148;
Best Local Similarity 66.7%; Pred. No. 2.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LYKWHG 7
Db 140 MYQWHG 145

Search completed: December 16, 2003, 14:41:05
Job time : 22.1667 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:35 ; Search time 5.66667 Seconds
(without alignments)
74.689 Million cell updates/sec

Title: US-09-870-089B-3

Perfect score: 57

Sequence: 1 FLYKWHGFV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	70.2	634	1 Y015 MYCPN	P75094 mycoplasma
2	39	68.4	819	1 ADVL_MOUSE	O88398 mus musculus
3	38	66.7	571	1 SYQ_EICAI	P57490 buchnera ap
4	38	66.7	589	1 Y015 MYCSE	P47261 mycoplasma
5	37	64.9	375	1 SWP_PIRAE	Q81u5 pyrobaculum
6	37	64.9	491	1 GABD_RHISN	P55653 rhizobium s
7	37	64.9	503	1 CCS_CITSI	Q9sea0 citrus sine
8	37	64.9	518	1 G6PD_DROYA	Q27638 drosophila
9	37	64.9	524	1 G6PD_DROME	P12646 drosophila
10	37	64.9	527	1 TP6B_PIRAE	Q8xvm0 pyrobaculum
11	36	63.2	153	1 LSPA_WIGBR	Q8d2r1 wiggleswort
12	36	63.2	405	1 GSFF_PSEAE	Q00513 pseudomonas
13	36	63.2	424	1 MTX2_XANOR	P52311 xanthomonas
14	36	63.2	485	1 SYE_STRMU	Q8dgv9 streptococc
15	36	63.2	486	1 SYE_STRPN	Q97ngl streptococc
16	36	63.2	486	1 SYE_STRB6	Q8cwn5 streptococc
17	36	63.2	578	1 Y0D6_CAERL	Q09266 caenorhabdi
18	36	63.2	628	1 HCYG_EURCA	Q9nfl4 eurytelma c
19	36	63.2	711	1 DCOE_ECOLI	P21169 escherichia
20	36	63.2	1584	1 YJ9G_YEAST	P47170 saccharomyc
21	35	61.4	66	1 YPJJ_ECOLI	P58033 escherichia
22	35	61.4	73	1 YKFX_ECOLI	Q9xb42 escherichia
23	35	61.4	164	1 LSPA_ECOS7	Q8xa48 escherichia
24	35	61.4	164	1 LSPA_ECOL6	Q8flb6 escherichia
25	35	61.4	164	1 LSPA_ECOL6	Q8flb6 escherichia
26	35	61.4	165	1 LSPA_ENTAE	P13514 enterobacte
27	35	61.4	193	1 COX2_MYTED	Q00227 mytilus edu
28	35	61.4	219	1 YQFA_ECOLI	Q46827 escherichia
29	35	61.4	351	1 MURB_BORPE	Q9x6y8 bordetella
30	35	61.4	530	1 MTM2_METUA	Q58843 methanococc
31	35	61.4	538	1 CP5D_CANNA	P16141 candida mal
32	35	61.4	626	1 HCY6_ANDAU	P80476 androctonus
33	35	61.4	626	1 HCYB_EURCA	Q9nfh9 eurytelma c

ALIGNMENTS

RESULT 1

Y015_MYCPN

ID Y015_MYCPN STANDARD; PRT; 634 AA.

AC P75094;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical ABC transporter ATP-binding protein MG015 homolog

DE (D12_016334).

GN MPN019 OR MP135.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=97105885; PubMed=8948633;

RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,

RA Hermann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

pneumoniae.";

RL Nucleic Acids Res. 24:4420-4449(1996).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MSBA SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AE000015; AAB95783.1; -.

DR PIR; S73461; S73461.

DR HSP; P13569; IINED.

DR InterPro; IPR003593; AAA_ATPase.

DR InterPro; IPR001140; ABC_TM_transpt.

DR InterPro; IPR003439; ABC_transporter.

DR Pfam; PF00664; ABC_membrane; 1.

DR Pfam; PF00005; ABC_tran; 1.

DR ProDom; PD000006; ABC_transporter; 1.

DR SMART; SM00382; AAA; 1.

DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.

DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.

DR Hypothetical protein; ATP-binding; Transport; Transmembrane;

KW Complete proteome.

FT TRANSMEM 54 74 POTENTIAL.

FT TRANSMEM 111 131 POTENTIAL.

FT TRANSMEM 189 209 POTENTIAL.

FT TRANSMEM 213 233 POTENTIAL.

FT TRANSMEM 296 316 POTENTIAL.

FT TRANSMEM 325 345 POTENTIAL.

FT NP_BIND 430 437 ATP (POTENTIAL).

SQ SEQUENCE 634 AA; 71147 MW; E10CE07E0EB9D406 CRC64;

```
Query Match      70.2%; Score 40; DB 1; Length 634;
Best Local Similarity 55.6%; Pred. No. 13;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 FLYKWHGFV 9
      ||| |||
Db      297 FLYSWFGFI 305

RESULT 2
ADVL MOUSE
ID ADVL_MOUSE STANDARD; PRT; 819 AA.
AC Q88398; O70466;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Advillin (p92) (Actin-binding protein DOC6).
GN AVIL OR ADVIL.
OS Mus musculus (Mouse).
OC Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=98330437; PubMed=9664034;
RA Marks P.W., Arai M., Bandura J.L., Kwiatkowski D.J.;
RT "Advillin (p92): a new member of the gelsolin/villin family of actin
regulatory proteins.";
RL J Cell Sci. 111:2129-2136(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH Swiss;
RX MEDLINE=98315054; PubMed=9649432;
RA Wang X.-Z., Kuroda M., Sok J., Batchvarova N., Kimmel R., Chung P.,
Zinszner H., Ron D.;
RT "Identification of novel stress-induced genes downstream of chop.";
RL EMBO J. 17:3619-3630(1998).
CC -1- FUNCTION: CA(2+)-REGULATED ACTIN-BINDING PROTEIN. MAY HAVE A
UNIQUE FUNCTION IN THE MORPHOGENESIS OF NEURONAL CELLS WHICH FORM
GANGLIA.
CC -1- TISSUE SPECIFICITY: MOST HIGHLY EXPRESSED IN THE ENDOMETRIUM OF
THE UTERUS, THE INTESTINAL VILLI AND THE TESTES. WEAKER EXPRESSION
ALSO DETECTED IN THE BRAIN AND ON THE SURFACE OF THE TONGUE.
CC -1- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION WAS DETECTED IN DORSAL
ROOT AND TRIGEMINAL GANGLIA BY EMBRYONIC DAY 14.5 AND CONTINUED
THROUGH AT LEAST DAY 16.5.
CC -1- SIMILARITY: BELONGS TO THE VILLIN/GELSOLIN FAMILY.
CC -1- SIMILARITY: Contains 6 gelsolin-like repeats.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; AF041448; AAC25050.1; -
DR EMBL; AF059486; AAC31808.1; -
DR HSP; P02640; 2VIL.
DR MGD; MG1.1333798; Avil.
DR GO; GO:0006950; P:response to stress; IDA.
DR InterPro; IPR001974; Gelsolin.
DR InterPro; IPR003128; VHP.
DR Pfam; PF00626; Gelsolin; 6.
DR PRINTS; PR00597; GELSOLIN.
DR SMART; SM00262; GEL; 6.
DR SMART; SM00153; VHP; 1.
DR Cytoskeleton; Calcium; Actin-binding; Capping protein; Repeat.
KW DOMAIN 1 731
FT DOMAIN 732 819
FT HEADPIECE (BY SIMILARITY).

Query Match      68.4%; Score 39; DB 1; Length 819;
Best Local Similarity 83.3%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      3 YKWHGF 8
      |:||||
Db      412 YQWHGF 417

RESULT 3
SYQ_BUCAI
ID SYQ_BUCAI STANDARD; PRT; 571 AA.
AC P57490;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
DE (Glns).
DE GLNS OR BU415.
GN Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbiotic bacterium).
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10933077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. ABS.";
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
diphosphate + L-glutaminyl-tRNA(Gln).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; AP001119; BAB13113.1; -
DR HSP; P00962; IGTR.
DR HAMAP; MF_00126; -; 1.
DR InterPro; IPR004514; Glns.
DR InterPro; IPR000924; Glu tRNA-synt_1c.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
```

DR TIGRFAMS; TIGR00440; glns; 1.
DR PROSITE; PS00178; AA TRNA-LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 35 45 "HIGH" REGION.
FT SITE 269 273 "KMSKS" REGION.
FT BINDING 272 272 ATP (BY SIMILARITY).
SQ SEQUENCE 571 AA; 67835 MW; 1FCE69932D563D52 CRC64;

Query Match 66.7%; Score 38; DB 1; Length 571;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 YKWHGFV 9
||| |
Db 92 YKWHGNV 98

RESULT 4

ID Y015 MYCGE STANDARD; PRT; 589 AA.
AC P47261;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein MG015.
GN MG015.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.P., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uitterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MSBA SUBFAMILY.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U39680; AAC71231.1; -.
DR PIR; F64201; F64201.
DR HSSP; P13569; 1NBD.
DR TIGR; MG015; -.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS0211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS0893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; Transport; Transmembrane;
KW Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 66 86 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 251 271 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.

FT TRANSMEM 303 323 POTENTIAL.
FT NP_BIND 385 392 ATP (POTENTIAL).
SQ SEQUENCE 589 AA; 66126 MW; 0EA6A816DD3A3CC9 CRC64;

Query Match 66.7%; Score 38; DB 1; Length 589;
Best Local Similarity 55.6%; Pred. No. 27;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLYKWHGFV 9
||| |
Db 252 FLFSWFGFI 260

RESULT 5

SYW_PYRAE STANDARD; PRT; 375 AA.
ID SYW_PYRAE
AC Q82TU5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase)
DE (TrpRS).
GN TRPS OR PAE3091.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
CC diphosphate + L-tryptophanyl-tRNA(Trp).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AE009911; AAL64664.1; -.
DR HAMAP; MF_00140; -; 1.
DR InterPro; IPR002305; tRNA-synt_1b.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002306; Trp tRNA-synt_1b.
DR Pfam; PF00579; tRNA-synt_1b; 1.
DR PRINTS; PR01039; TRNASYNTHTRP.
DR TIGRFAMS; TIGR00233; trps; 1.
DR PROSITE; PS00178; AA TRNA_LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 81 89 "HIGH" REGION.
FT SITE 258 262 "KMSKS" REGION.
SQ SEQUENCE 375 AA; 43178 MW; 3DDBF85DA680F116 CRC64;

Query Match 64.9%; Score 37; DB 1; Length 375;
Best Local Similarity 71.4%; Pred. No. 27;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLYKWHG 7
|: |||
Db 62 FIMKWHG 68

```

RESULT 6
GABD RHISN          STANDARD;          PRT;    491 AA.
ID  GABD RHISN
AC  P55653;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Probable succinate-semialdehyde dehydrogenase [NADP+] (EC 1.2.1.16)
DE  (SSDH).
DE  GABD OR Y4SJ.
GN  Rhizobium sp. (strain NGR234).
OS  Plasmid sym pNGR234a.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC  Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX  NCBI_TaxID=394;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=97305956; PubMed=9163424;
RA  Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA  Perret X.;
RT  "Molecular basis of symbiosis between Rhizobium and legumes.";
RL  Nature 387:394-401(1997).
CC  -1- CATALYTIC ACTIVITY: Succinate semialdehyde + NAD(P)(+) + H(2)O =
CC  succinate + NAD(P)H.
CC  -1- PATHWAY: 4-aminobutyrate (GABA) degradation.
CC  -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
EMBL; AE000096; AAB91849.1; -.
DR  EMBL; S43963; S43963.
DR  HSP; P05091; ICW3.
DR  InterPro; IPR002086; Aldehyde_dehydr.
DR  Pfam; PF00171; aldedh; 1.
DR  PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; FALSE_NEG.
DR  PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW  Oxidoreductase; NADP; Plasmid.
FT  NP_BIND 241 246 NADP (ADP PART) (BY SIMILARITY).
FT  ACT_SITE 263 263 BY SIMILARITY.
FT  ACT_SITE 297 297 BY SIMILARITY.
SQ  SEQUENCE 491 AA; 53253 MW; FC54C5DF7B4D1B14 CRC64;

Query Match 64.9%; Score 37; DB 1; Length 491;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LYKWHGFV 9
Db 83 LWKWHRFI 90

RESULT 7
CCS_CITSI
AC  Q9SEAO;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Capsanthin/capsorubin synthase, chloroplast precursor.
DE  CCS.
GN  Citrus sinensis (Sweet orange).
OS  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC  eurosids II; Sapindales; Rutaceae; Citrus.
OX  NCBI_TaxID=2711;
RN  [1]
RP  SEQUENCE FROM N.A.

```

```

RA Xu C.J., Chen D.M., Zhang S.L.;
RT "Molecular cloning of capsanthin/capsorubin synthase gene from orange
RT (Citrus sinensis).";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE CONVERSION OF THE UBIQUITOUS 5,6-
CC EPOXYCAROTENOIDS, ANTERAXANTHIN AND VIOLAXANTHIN, INTO CAPSANTHIN
CC AND CAPSORUBIN, RESPECTIVELY.
CC -1- PATHWAY: Carotenoid biosynthesis.
CC -1- SUBCELLULAR LOCATION: Chloroplast; chloroplast (By similarity).
CC -1- SIMILARITY: BELONGS TO THE LYCOPENE CYCLASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; AF169241; AAF18389.1; -.
DR  EMBL; AF169241; AAF18389.1; -.
KW  Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
KW  Transit peptide.
FT  TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT  CHAIN ? 503 CAPSANTHIN/CAPSORUBIN SYNTHASE.
FT  NP_BIND 88 117 NAD (POTENTIAL).
SQ  SEQUENCE 503 AA; 56663 MW; 2C957FOAA91075B6 CRC64;

Query Match 64.9%; Score 37; DB 1; Length 503;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 YKWHGFV 9
Db 449 YNWHGFL 455

RESULT 8
G6PD DROYA          STANDARD;          PRT;    518 AA.
ID  G6PD DROYA
AC  Q27638;
DT  15-JUL-1998 (Rel. 36, Created)
DT  15-JUL-1998 (Rel. 36, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD) (Fragment).
DE  ZW OR G6PD.
OS  Drosophila yakuba (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7245;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BG1016;
RA  Eanes W.F., Kirchner M., Yoon J., Biermann C., Wang I., McCartney M.,
RA  Verrelli B.C.;
RL  Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
CC 1,5-lactone 6-phosphate + NADPH.
CC -1- PATHWAY: Pentose phosphate pathway; first step.
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; U42750; AAB02813.1; -.
DR  EMBL; U42750; AAB02813.1; -.
DR  HSP; P11411; LDPG.
DR  FlyBase; FBgn0015691; Dyak\Zw.

```

```

DR InterPro: IPR001282; G6PD.
DR Pfam: PF00479; G6PD; 1.
DR Pfam: PF02781; G6PD_C; 1.
DR PRINTS: PR00079; G6PDHGRNASE.
DR ProDom: PD001129; G6PD; 1.
DR TIGRFAMs: TIGR00871; zwf; 1.
DR PROSITE: PS0069; G6P DEHYDROGENASE; 1.
KW Oxidoreductase; NADP; Glucose metabolism.
FT NON_TER 1
FT ACT_SITE 203 BY SIMILARITY.
SQ SEQUENCE 518 AA; 59938 MW; 6AFA194E1D3A3E8B CRC64;

Query Match 64.9%; Score 37; DB 1; Length 518;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YKWHG 7
Db 506 YKWHG 510
|||||

RESULT 9
ID G6PD DROME STANDARD; PRT; 524 AA.
AC P12646; Q27574; Q27872; Q27879; Q27881; Q9VWE2; Q9VWE3;
DT 01-OCT-1989 (Rel. 12, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD)
DE (Zwischenferment).
GN ZW OR G6PD OR CGI2529.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=89255872; PubMed=2838391;
RC Fouts D., Ganguly R., Gutierrez A.G., Lucchesi J.C., Manning J.E.;
RT "Nucleotide sequence of the Drosophila glucose-6-phosphate
RL dehydrogenase gene and comparison with the homologous human gene.";
RN Gene 63:261-275(1988).
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=F23.3, F24.1, M741, MT68, Z3, Z5, Z11, Z16, Z21, Z27, Z41, Z42,
RC Z55, Z62, Z64, and Z74;
RX MEDLINE=97070821; PubMed=8913747;
RA Eanes W.F., Kirchner M., Yoon J., Biermann C.H., Wang I.N.,
RA McCartney M.A., Verrelli B.C.;
RT "Historical selection, amino acid polymorphism and lineage-specific
RT divergence at the G6pd locus in Drosophila melanogaster and D.
RT simulans.";
RL Genetics 144:1027-1041(1996).
[3]
RN RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Brokstein P., Brottier P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

```

```

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphry L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusse K.D., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -!- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
CC 1,5-lactone 6-phosphate + NADPH.
CC -!- PATHWAY: Pentose phosphate pathway; first step.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P12646-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P12646-2; Sequence=VSP 001593;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

EMBL; M26674; AA51463.1; -
EMBL; M26673; AA51463.1; JOINED.
EMBL; U42738; AAB02801.1; -
EMBL; U42739; AAB02802.1; -
EMBL; U42740; AAB02803.1; -
EMBL; U42741; AAB02804.1; -
EMBL; U42742; AAB02805.1; -
EMBL; U42743; AAB02806.1; -
EMBL; U42744; AAB02807.1; -
EMBL; U42745; AAB02808.1; -
EMBL; U42746; AAB02809.1; -
EMBL; U42747; AAB02810.1; -
EMBL; U42748; AAB02811.1; -
EMBL; U42749; AAB02812.1; -
EMBL; U43165; AAA99071.1; -
EMBL; U43166; AAA99072.1; -
EMBL; U43167; AAA99073.1; -
EMBL; U44721; AAA99092.1; -
EMBL; U45985; AAA99107.1; -
EMBL; AB003512; AAF48999.1; -

```

```

DR EMBL; AF003512; AAF49000.1; -.
DR EMBL; AF052079; AAK93503.1; -.
DR PIR; A47740; A47740.
DR PIR; J02072; DEFFG6.
DR HSSP; P11411; 1DPG.
DR FlyBase; FBgn004057; Zw.
DR InterPro; IPR001282; G6PD.
DR Pfam; PF00479; G6PD; 1.
DR Pfam; PF02781; G6PD; 1.
DR PRINTS; PR00079; G6PDHGRNASE.
DR ProDom; PD001129; G6PD; 1.
DR TIGRfams; TIGR00871; zwf; 1.
DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
KW Oxidoreductase; NADP; Glucose metabolism; Polymorphism;
KW Alternative splicing.
FT ACT SITE 209 209
FT VARSPLTIC 1 7
FT FT
FT VARIANT 32 32
FT VARIANT 80 80
FT VARIANT 382 382
FT FT
FT FT
FT CONFLICT 185 186
FT CONFLICT 214 214
FT CONFLICT 344 346
FT CONFLICT 461 465
SQ SEQUENCE 524 AA; 60431 MW; AFB81B763A82F1FD CRC64;

Query Match 64.9%; Score 37; DB 1; Length 524;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 YKWHG 7
Db 512 YKWHG 516

RESULT 10
TP6B PYRAE STANDARD; PRT; 527 AA.
AC Q8ZVM0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Type II DNA topoisomerase VI subunit B (EC 5.99.1.3) (TopoVI-B).
GN TOP6B OR PAE2217.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.
"Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum.";
RT aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -!- FUNCTION: Relaxes both positive and negative superturns and
CC exhibits a strong decatenase activity. The B subunit binds ATP (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: Heterotetramer of two subunits A and two subunits B (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE TOP6B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AF063521; BAC24439.1; -.
DR HAMAP; MF 00161; -.
DR InterPro; IPR001872; SigPase A8.
DR Pfam; PF01252; Peptidase A8; 1.
DR PRINTS; PR00781; LIPOSIGTASE.
DR ProDom; PD004304; SigPase A8; 1.
DR TIGRfams; TIGR00077; lspa-1.
DR PROSITE; PS00855; SPASE-II; FALSE_NEG.
KW Hydrolase; Aspartyl protease; Transmembrane; Inner membrane;

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AF009863; AAL64036.1; -.
DR HAMAP; MF 00322; -.
DR InterPro; IPR003594; ATPbind ATPase.
DR InterPro; IPR005734; DNA_top6B.
DR Pfam; PF02518; HATPase_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR TIGRfams; TIGR01052; top6b; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Complete proteome.
SQ SEQUENCE 527 AA; 59870 MW; 7BAADB81F8FC9D97 CRC64;

Query Match 64.9%; Score 37; DB 1; Length 527;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 YKWHG 7
Db 162 YKWHG 166

RESULT 11
LSPA WIGBR
ID LSPA WIGBR STANDARD; PRT; 153 AA.
AC Q8D2R1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lipoprotein signal peptidase (EC 3.4.23.36) (Prolipoprotein signal
DE peptidase) (Signal peptidase II) (SPase II).
GN LSPA OR WIGBR2930.
OS Wigglesworthia glossinidia brevipalpis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Wigglesworthia.
OX NCBI_TaxID=36870;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22297718; PubMed=12219091;
RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Aksoy S.;
RT "Genome sequence of the endocellular obligate symbiont of tsetse
RT flies, Wigglesworthia glossinidia.";
RL Nat. Genet. 32:402-407(2002).
CC -!- FUNCTION: This protein specifically catalyzes the removal of
CC signal peptides from prolipoproteins (By similarity).
CC -!- CATALYTIC ACTIVITY: Cleavage of N-terminal leader sequences from
CC membrane prolipoproteins. Hydrolyses Xaa-Xbb-Xcc-|-Cys, in which
CC Xaa is hydrophobic (preferably Leu), Xbb is often Ser or Ala, Xcc
CC is often Gly or Ala, and the Cys is alkylated on sulfur with a
CC diacylglyceryl group.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -!- SIMILARITY: Belongs to peptidase family A8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AB063521; BAC24439.1; -.
DR HAMAP; MF 00161; -.
DR InterPro; IPR001872; SigPase A8.
DR Pfam; PF01252; Peptidase A8; 1.
DR PRINTS; PR00781; LIPOSIGTASE.
DR ProDom; PD004304; SigPase A8; 1.
DR TIGRfams; TIGR00077; lspa-1.
DR PROSITE; PS00855; SPASE-II; FALSE_NEG.
KW Hydrolase; Aspartyl protease; Transmembrane; Inner membrane;

```

KW Complete proteome. 7 24 Potential.
 FT TRANSMEM 60 79 Potential.
 FT TRANSMEM 86 108 Potential.
 FT TRANSMEM 123 145 Potential.
 FT ACT_SITE 105 105 BY SIMILARITY.
 FT ACT_SITE 132 132 BY SIMILARITY.
 SQ SEQUENCE 153 AA; 18085 MW; 1AEE3B201E935B58 CRC64;

Query Match 63.2%; Score 36; DB 1; Length 153;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLYKWH 6
 :|||||
 Db 119 YIVKWH 124

RESULT 12
 GSPP PSEAE
 ID GSPP PSEAE STANDARD; PRT; 405 AA.
 AC Q00513;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE General secretion pathway protein F.
 GN XCP5 OR PA3102.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=92269572; PubMed=1588814;
 RA Bally M., Filloux A., Akrim M., Ball G., Lazdunski A., Tommassen J.;
 RT "Protein secretion in Pseudomonas aeruginosa: characterization of
 seven xcp genes and processing of secretory apparatus components by
 RT prelipin peptidase.";
 RL Mol. Microbiol. 6:1121-1131(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huynh W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
 CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF A VARIETY
 CC OF ENZYMES ACROSS THE OUTER MEMBRANE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -!- SIMILARITY: BELONGS TO THE PULF/OUTF/EXEF/XPSF/XPCS FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X62666; CAA4534.1; -.
 CC EMBL; AE004734; AAG06490.1; -.
 CC PIR; S25385; SKPSXS.
 CC InterPro; IPR003004; Bac GSPP.
 CC InterPro; IPR001992; Bact_secr_systII.

DR Pfam; PF00482; GSPF1 F; 1.
 DR PRINTS; PR00812; BCTERIALGSPF.
 DR PROSITE; PS00874; T2SP F; 1.
 KW Transport; Transmembrane; Inner membrane; Complete proteome.
 FT TRANSMEM 169 189 POTENTIAL.
 FT TRANSMEM 219 239 POTENTIAL.
 FT ACT_SITE 377 397 POTENTIAL.
 SQ SEQUENCE 405 AA; 44061 MW; BEE59B58724C167E CRC64;

Query Match 63.2%; Score 36; DB 1; Length 405;
 Best Local Similarity 55.6%; Pred. No. 42;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLYKWHGFV 9
 :|||||
 Db 248 FRERWGHFL 256

RESULT 13
 MTX2 XANOR
 ID MTX2 XANOR STANDARD; PRT; 424 AA.
 AC P52311;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Modification methylase XorII (EC 2.1.1.73) (Cytosine-specific
 DE methyltransferase XorII) (M.XorII).
 GN XORIIM.
 OS Xanthomonas oryzae (pv. oryzae).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=64187;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JW89011;
 RX MEDLINE=94359461; PubMed=8078464;
 RA Choi S.H., Leach J.E.;
 RT "Identification of the XorII methyltransferase gene and a var homolog
 RT from Xanthomonas oryzae pv. oryzae.";
 RL Mol. Gen. Genet. 244:383-390(1994).
 CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
 CC CGATCG. CAUSES SPECIFIC METHYLATION ON C-? ON BOTH STRANDS AND
 CC PROTECTS THE DNA FROM CLEAVAGE BY THE XORI methyltransferase.
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
 CC adenosyl-L-homocysteine + DNA 5-methylcytosine.
 CC -!- SIMILARITY: BELONGS TO THE CS-METHYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; U06424; AAA50432.1; -.
 CC PIR; S46293; S46293.
 CC HSSP; P20589; 1DCT.
 CC REBASE; 3534; M.XorII.
 CC InterPro; IPR001525; CS_DNA_meth.
 CC Pfam; PF00145; DNA_methylase; 1.
 CC PRINTS; PR00105; CS_METHYLFRASE.
 CC TIGRFAMs; TIGR00675; dcm; 1.
 CC PROSITE; PS00094; C5_MTASE_1; 1.
 CC PROSITE; PS00095; C5_MTASE_2; 1.
 KW Transferase; Methyltransferase; Restriction system.
 FT ACT_SITE 83 83 BY SIMILARITY.
 SQ SEQUENCE 424 AA; 46971 MW; F2EEB12F09918781 CRC64;

Query Match 63.2%; Score 36; DB 1; Length 424;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

QY      4 KWHGF 8
Db      337 KWHGF 341

RESULT 14
SYE_STRMU
ID_SYE_STRMU STANDARD; PRT; 485 AA.
AC Q8DVX9;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
DE (GluRS).
GN GLTX OR SMU.330.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014881; AAN58089.1; -.
CC HAMAP; MF_00022; -. 1 GltX bact.
CC InterPro; IPR004527; GltX bact.
CC InterPro; IPR000924; Glu tRNA-synt_1c.
CC InterPro; IPR001412; tRNA-synt_1.
CC Pfam; PF00749; tRNA-synt_1c; 1.
CC PRINTS; PR00987; TRNASYNTHGLU.
CC TIGRfams; TIGR00464; GltX bact; 1.
CC PROSITE; PS00178; AA tRNA_LIGASE_I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 11 21 "HIGH" REGION.
CC SITE 255 259 "KMSKS" REGION.
CC BINDING 258 258 ATP (BY SIMILARITY).
CC SEQUENCE 485 AA; 55720 MW; DFA1B99DDCD1363F CRC64;

Query Match 63.2%; Score 36; DB 1; Length 485;
Best Local Similarity 62.5%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      2 LYKWHGFV 9
Db      165 IYKWHDMV 172

RESULT 15
SYE_STRPN
ID_SYE_STRPN STANDARD; PRT; 486 AA.
AC Q97NGI;
DT 15-SEP-2003 (Rel. 42, Created)

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:09:45 ; Search time 25 seconds
(without alignments)

92.899 Million cell updates/sec

Title: US-09-870-089B-3

Perfect score: 57

Sequence: 1 FLYKWHGFV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	75.4	690	5	Q27598 drosophila
2	43	75.4	690	5	Q3V896 drosophila
3	43	75.4	691	5	Q3V083 sarcophaga
4	42	73.7	685	5	O44249 manduca sex
5	41	71.9	144	10	Q9AW53 guillardi
6	41	71.9	459	16	Q8X115 drosophila
7	41	71.9	1149	5	Q81LP5 plasmodium
8	40	70.2	683	5	Q95R43 drosophila
9	40	70.2	683	5	Q9W1V6 drosophila
10	40	70.2	683	5	Q81IF6 drosophila
11	40	70.2	693	5	Q9BLG6 drosophila
12	40	70.2	693	5	Q9GU90 bombyx mori
13	40	70.2	693	5	Q27452 bombyx mori
14	40	70.2	697	5	Q76208 hyphantria
15	39	68.4	268	16	Q8R946 thermoplasma
16	39	68.4	431	5	Q818D7 entamoeba h

17	39	68.4	805	2	Q9RGZ5
18	39	68.4	805	16	Q8ELV6
19	39	68.4	829	11	Q9WU06
20	38	66.7	291	16	Q8ZD70
21	38	66.7	446	17	Q97U17
22	38	66.7	448	10	Q9LJA6
23	38	66.7	653	5	Q95V11
24	38	66.7	684	5	Q816K2
25	38	66.7	685	5	Q9Y0B4
26	38	66.7	804	16	P70811
27	37	64.9	166	16	Q8D404
28	37	64.9	301	3	Q12350
29	37	64.9	382	12	O72591
30	37	64.9	400	6	Q95JT8
31	37	64.9	406	4	Q8WY11
32	37	64.9	552	16	O92AL0
33	37	64.9	552	16	O8Y698
34	37	64.9	670	5	Q9BLD9
35	37	64.9	684	5	Q9V521
36	37	64.9	687	5	Q8WS15
37	37	64.9	688	5	O44250
38	37	64.9	690	5	Q9GVA6
39	37	64.9	699	5	Q9GVA5
40	37	64.9	777	16	Q8CMW9
41	36	63.2	153	16	Q8D2R1
42	36	63.2	225	17	Q9HPL6
43	36	63.2	242	16	Q9CMR9
44	36	63.2	243	11	Q9CTR2
45	36	63.2	295	11	Q99N31

ALIGNMENTS

RESULT 1

ID	Q27598	PRELIMINARY;	PRT;	690 AA.
AC	Q27598			
DT	01-NOV-1996	(TrEMBLrel. 01, Created)		
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)		
DT	01-OCT-2002	(TrEMBLrel. 22, Last annotation update)		
DE	PRO-phenol oxidase A1 (EC 1.14.18.1) (BLACK cells protein).			
GN	BC OR CG5779.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=OREGON-R;			
RX	MEDLINE=95372361; PubMed=7644493;			
RA	Fujimoto K., Okino N., Kawabata S., Iwanaga S., Ohnishi E.;			
RT	Nucleotide sequence of the cDNA encoding the proenzyme of phenol oxidase A1 of Drosophila melanogaster.			
RL	Proc. Natl. Acad. Sci. U.S.A. 92:7769-7773(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RA	Ceiniker S.E., Aghavani A., Arcaina T.T., Baxter E., Blazej R.G.,			
RA	Chew M., Doyle C.M., Farfan D.E., Flanagan J., Houston K.A.,			
RA	Humastti S.R., Karra K., Kearney L., Kim S.H., Lee B., Lomotan M.A.,			
RA	Mazda P., Mok M.S., Moshrefi A.R., Moshrefi M., Nixon K., Pacleb J.M.,			
RA	Park S., Pfeiffer B., Punch D., Snir E., Twomey B., Wan K.H.,			
RA	Whitelaw K.R., Yee A., Zhang R., Zieran L.L., Kimmel B.;			
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; D45835; BAA08272.1; -			
DR	HSSP; P04253; 1LLA.			
DR	FlyBase; FBgn0000165; Bc.			
DR	InterPro; IPR000896; Hemocyanin.			
DR	InterPro; IPR005203; hemocyanin_C.			
DR	InterPro; IPR005204; hemocyanin_N.			

Q9rgz5	bacillus ps
Q8elv6	oceanobacil
Q9wu06	rattus norv
Q8zdt0	versinia pe
Q97uy7	sulfolobus
Q91jac	arabidopsis
Q95vll	asterias ru
Q816k2	holotrichia
Q9y0b4	sarcophaga
P70811	bacillus ha
Q8d404	vibrio vuln
Q12350	saccharomyc
O72591	beet soli-b
Q95jt8	macaca fasc
Q8wy11	homo sapien
Q92al0	listeria in
Q8y698	listeria mo
Q9bld9	drosophila
Q9v521	drosophila
Q8ws15	aedes aegypt
O44250	anopheles g
Q9gva6	pimpla hypo
Q9gva5	pimpla hypo
Q8cmw9	staphylococ
Q8d2r1	wiggleswort
Q9hpl6	haemobacteri
Q9cmr9	pasteurella
Q9ctr2	mus musculu
Q99n31	mus musculu

```

DR InterPro: IPR002227; Tyrosinase.
DR Pfam: PF00372; hemocyanin; 1.
DR Pfam: PF03723; hemocyanin C; 1.
DR Pfam: PF03722; hemocyanin N; 1.
DR PRINTS: PR00187; HAEMOCYANIN.
DR PROSITE: PS00209; HEMOCYANIN 1; 1.
DR PROSITE: PS00210; HEMOCYANIN 2; 1.
DR PROSITE: PS00498; TYROSINASE 2; 1.
KW Oxidoreductase; Polymorphism.
FT VARIANT 387 387 D -> E (IN STRAIN BERKELEY).
FT VARIANT 605 605 L -> V (IN STRAIN BERKELEY).
SQ SEQUENCE 690 AA; 79090 MW; 7F2D6564651B1F58 CRC64;

Query Match 75.4%; Score 43; DB 5; Length 690;
Best Local Similarity 71.4%; Pred. No. 36;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 YKWHGFV 9
Db 404 YRWHGFI 410

RESULT 2
Q9V896 PRELIMINARY; PRT; 690 AA.
ID Q9V896
AC Q9V896;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE BC gene product (GH04080P).
GN BC OR CG5779.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan X.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.D.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherly J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M.C., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

```

```

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AR003801; AAF57775.1; -.
DR EMBL: AY060617; AAL28165.1; -.
DR HSSP: P04253; 1LLA.
DR FlyBase; FBgn0000165; Bc.
DR InterPro: IPR000896; Hemocyanin.
DR InterPro: IPR005203; hemocyanin C.
DR InterPro: IPR005204; hemocyanin N.
DR InterPro: IPR002227; Tyrosinase.
DR Pfam: PF00372; hemocyanin; 1.
DR Pfam: PF03723; hemocyanin C; 1.
DR Pfam: PF03722; hemocyanin N; 1.
DR PRINTS: PR00187; HAEMOCYANIN.
DR PROSITE: PS00209; HEMOCYANIN 1; 1.
DR PROSITE: PS00210; HEMOCYANIN 2; 1.
DR PROSITE: PS00498; TYROSINASE 2; 1.
SQ SEQUENCE 690 AA; 79090 MW; 49FC131F1D08B893 CRC64;

Query Match 75.4%; Score 43; DB 5; Length 690;
Best Local Similarity 71.4%; Pred. No. 36;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 YKWHGFV 9
Db 404 YRWHGFI 410

RESULT 3
Q9Y0B3 PRELIMINARY; PRT; 691 AA.
ID Q9Y0B3
AC Q9Y0B3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Phenoloxidase.
GN PPO2.
OS Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
OC Sarcophagidae; Sarcophaga.
OX NCBI_TaxID=7385;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20359206; PubMed=10899462;
RA Chase M.R., Raina K., Bruno J., Sugumaran M.;
RT "Purification, characterization and molecular cloning of
RT phenoloxidases from Sarcophaga bullata.";
RL Insect Biochem. Mol. Biol. 30:953-967(2000).
DR EMBL: AF161261; AAD45527.1; -.
DR HSSP: P04254; 1HC2.
DR InterPro: IPR000896; Hemocyanin.
DR InterPro: IPR005203; hemocyanin C.
DR InterPro: IPR005204; hemocyanin N.
DR InterPro: IPR00504; RNA_rec_mot.
DR InterPro: IPR002227; Tyrosinase.
DR Pfam: PF00372; hemocyanin; 1.
DR Pfam: PF03723; hemocyanin C; 1.
DR Pfam: PF03722; hemocyanin N; 1.
DR PRINTS: PR00187; HAEMOCYANIN.
DR PROSITE: PS00209; HEMOCYANIN 1; 1.

```

```

DR PROSITE; PS00210; HEMOCYANIN 2; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
SQ SEQUENCE 691 AA; 79796 MW; DD754582CACBC94D CRC64;

Query Match 75.4%; Score 43; DB 5; Length 691;
Best Local Similarity 71.4%; Pred. No. 36;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 YKWHGFV 9
Db 404 YRWKGF 410

RESULT 4
O44249 PRELIMINARY; PRT; 685 AA.
ID O44249
AC O44249
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pro-phenol oxidase subunit 1 (EC 1.14.18.1).
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; SpHINGIODEA;
OC SpHINGIDAE; SpHINGINAE; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98135121; PubMed=9474780;
RA Jiang H., Wang Y., Ma C., Kanost M.R.;
RT "Subunit composition of pro-phenol oxidase from Manduca sexta:
RT molecular cloning of subunit PROPO-PI."
RL Insect Biochem. Mol. Biol. 27:835-850(1997).
DR EMBL; AF003253; AAC05796.1; -.
DR HSP; P04253; IOXY.
DR InterPro; IPR00896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR Pfam; PF002227; Tyrosinase.
DR Pfam; PF0372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
KW Oxidoreductase.
SQ SEQUENCE 685 AA; 78965 MW; BE5811D145302583 CRC64;

Query Match 73.7%; Score 42; DB 5; Length 685;
Best Local Similarity 66.7%; Pred. No. 52;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGFV 9
Db 401 FFYKWHGFV 409

RESULT 5
Q9AW53 PRELIMINARY; PRT; 144 AA.
ID Q9AW53
AC Q9AW53
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Ubiquitin-conjugating enzyme E2-21 KD (EC 6.3.2.19) (Ubiquitin-protein
DE ligase) (Ubiquitin carrier protein).
DE UCE-E2.
GN UCE-E2.
OS Guillardia theta (Cryptomonas phi).
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=20087226; PubMed=10618395;
RA Zauner S., Fraunholz M., Wastl J., Penny S.L., Beaton M.,
RA Cavalier-Smith T., Maier U., Douglas S.;
RT "Chloroplast protein and centrosomal genes, a tRNA intron, and odd
RT telomeres in an unusually compact eukaryotic genome, the cryptomonad
RT nucleomorph.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2123349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.;
RT "The highly reduced genome of an enslaved algal nucleus.";
RL Nature 410:1091-1096(2001).
CC -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC -!- PROTEINS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +
CC -!- DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.
CC -!- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.
CC -!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC -!- THIOLESTER FORMATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR EMBL; AJ010592; CAC27017.1; -.
DR HSP; P15731; IQCQ.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCG; 1.
DR PROSITE; PS00183; UBIQUITIN CONJUGAT_1; 1.
DR PROSITE; PS00127; UBIQUITIN CONJUGAT_2; 1.
KW Ligase; Ubl conjugation pathway.
SQ SEQUENCE 144 AA; 16365 MW; D8D4B2DD6EF74E80 CRC64;

Query Match 71.9%; Score 41; DB 10; Length 144;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYKWHGFV 9
Db 31 LYKWKGF 38

RESULT 6
Q8X115 PRELIMINARY; PRT; 459 AA.
ID Q8X115
AC Q8X115;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-glucosidase.
DE CPE2309.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AF003193; BABG2015.1; -.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDLASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
KW Complete proteome.
SQ SEQUENCE 459 AA; 53502 MW; FD19813255CA2686 CRC64;

```

```

Query Match          71.9%; Score 41; DB 16; Length 459;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLYKWH 6
Db 175 FLYKWH 180

RESULT 7
Q8ILP5 PRELIMINARY; PRT; 1149 AA.
AC Q8ILP5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GS PF14 0199.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perlea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511 (2002).
DR EMBL; AE014819; RAN36811.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1149 AA; 137812 MW; 5C3B42A34DCBFCDC CRC64;

Query Match          71.9%; Score 41; DB 5; Length 1149;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGFV 9
Db 869 FLYKWHGFV 877

RESULT 8
Q95R43 PRELIMINARY; PRT; 683 AA.
AC Q95R43;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE SD09866P.
GS DOX-A3 OR CG2952.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061624; AAL29172.1; -.
FlyBase; FBgn0000487; Dox-A3.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin C.
DR InterPro; IPR005204; hemocyanin N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin C; 1.
DR Pfam; PF03722; hemocyanin N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN 1; 1.
DR PROSITE; PS00210; HEMOCYANIN 2; 1.
DR PROSITE; PS00498; TYROSINASE 2; 1.
SQ SEQUENCE 683 AA; 79251 MW; E0B3F7E32B40F20B CRC64;

Query Match          70.2%; Score 40; DB 5; Length 683;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 YKWHGFV 9
Db 403 YKWHAFI 409

RESULT 9
Q9WIV6 PRELIMINARY; PRT; 683 AA.
AC Q9WIV6;
DT 01-WAY-2000 (TREMBLrel. 13, Created)
DT 01-WAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE CG2952 protein.
GS DOX-A3 OR CG2952.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.D.G.,
RA Abil J.F., Agbayani A., An H.-J., Andrews-Fianknoch C., Baldwin D.,
RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dudson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidgen-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs K.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003459; AAF46946.1; -.
DR HSP; P04253; 10XY.
DR FLYBase; FBgn0000487; Ddx-A3.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
SQ SEQUENCE 683 AA; 79314 MW; 437CBDD9E8A278BF CRC64;

Query Match 70.2%; Score 40; DB 5; Length 683;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 YKWHGFV 9
Db 403 YKWHAFI 409

RESULT 10
ID Q811F6 PRELIMINARY; PRT; 683 AA.
AC Q811F6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE DDX-A3-PA.
GN DDX-A3.
OS Drosophila erecta (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tucson 14021-0224.0;
RX MEDLINE=22426072; PubMed=12537575;
RA Bergman C.M., Pfeiffer B.D., Rincon-Limas D.E., Hoskins R.A.,
RA Gierke A., Mungall C.J., Wang A.M., Kronmiller B., Pacלב J., Park S.,
RA Stapleton M., Wan K., George R.A., de Jong P.J., Botas J., Rubin G.M.,
RA Celniker S.E.;
RT "Assessing the impact of comparative genomic sequence data on the
RT functional annotation of the Drosophila genome.";
RL Genome Biol. 3:RESEARCH0086.1-RESEARCH0086.20(2002).
DR EMBL; AY190941; AA001013.1; -.
SQ SEQUENCE 683 AA; 79054 MW; 9FF9A0E2A970662A CRC64;

Query Match 70.2%; Score 40; DB 5; Length 683;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 YKWHGFV 9
Db 403 YKWHAFI 409

RESULT 11
ID Q9BLG6 PRELIMINARY; PRT; 693 AA.
AC Q9BLG6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

```

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Prophenoloxidase-2s.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179161; PubMed=11118441;
RA Asano T., Ashida M.;
RT "Cuticular pro-phenoloxidase of the silkworm, Bombyx mori.
RT Purification and demonstration of its transport from hemolymph.";
RL J. Biol. Chem. 276:11100-11112(2001).
DR EMBL; AB048761; BAB40959.1; -.
DR HSP; P04253; 10XY.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
SQ SEQUENCE 693 AA; 80181 MW; 3A7D70BC5AC1720A CRC64;

Query Match 70.2%; Score 40; DB 5; Length 693;
Best Local Similarity 44.4%; Pred. No. 1.1e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLYKWHGFV 9
Db 401 FFYRNHAYI 409

RESULT 12
Q9GU90
ID Q9GU90 PRELIMINARY; PRT; 693 AA.
AC Q9GU90;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Prophenoloxidase-2 (EC 1.14.18.1).
GN PRO-PO2.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=a80;
RA Yamamoto K., Yakiyama M., Fujii H., Kusakabe T., Koga K., Aso Y.,
RA Ishiguro M.;
RT "cDNA Cloning and Overexpression of Prophenoloxidase from the Silkworm
RT Bombyx mori (a80 strain) and Developmental Changes of
RT Prophenoloxidase.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF178461; AAG09303.1; -.
DR HSP; P04253; 10XY.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.

```

DR PROSITE; PS00498; TYROSINASE_2; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 693 AA; 79994 MW; FD71A7593A9312EB CRC64;
 Query Match 70.2%; Score 40; DB 5; Length 693;
 Best Local Similarity 44.4%; Pred. No. 1.1e+02;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGFV 9
 | | | | : :
 Db 401 FFYRWHAYI 409

RESULT 13
 Q27452 Q27452 PRELIMINARY; PRT; 693 AA.
 AC Q27452;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE PROPHENOXIDASE subunit 2 (BC 1.14.18.1) (PROPHENOXIDASE-2F).
 OS Bombyx mori (Silk moth)
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
 OC Bombycidae; Bombyx.
 OX NCBI_TaxID=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KINSHU X SHOWA; TISSUE=Hemocyte;
 RX MEDLINE=95372362; PubMed=7644494;
 RA Kawabata T., Yasuhara Y., Ochiai M., Matsuura S., Ashida M.;
 RT "Molecular cloning of insect pro-phenol oxidase: a copper-containing
 RT protein homologous to arthropod hemocyanin."
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7774-7778(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Asano T., Ashida M.;
 RT "Pro-phenoloxidase type-2(f) 2.3 kbp Bombyx mori."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D49371; BAA08369.1; -.
 DR EMBL; AB048762; BAB41101.1; -.
 DR HSP; P04253; IOXY.
 DR InterPro; IPR000896; Hemocyanin.
 DR InterPro; IPR005203; hemocyanin_C.
 DR InterPro; IPR005204; hemocyanin_N.
 DR InterPro; IPR002227; Tyrosinase.
 DR Pfam; PF00372; hemocyanin_1.
 DR Pfam; PF03723; hemocyanin_C; 1.
 DR Pfam; PF03722; hemocyanin_N; 1.
 DR PRINTS; PR00187; HAEMOCYANIN.
 DR PROSITE; PS00209; HEMOCYANIN_1; 1.
 DR PROSITE; PS00210; HEMOCYANIN_2; 1.
 DR PROSITE; PS00498; TYROSINASE_2; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 693 AA; 80119 MW; 013114DF4CC1A926 CRC64;

Query Match 70.2%; Score 40; DB 5; Length 693;
 Best Local Similarity 44.4%; Pred. No. 1.1e+02;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGFV 9
 | | | | : :
 Db 401 FFYRWHAYI 409

RESULT 14
 O76208 O76208 PRELIMINARY; PRT; 697 AA.
 AC O76208;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Prophenoloxidase.

GN PROPO-2.
 OS Hyphantria cunea (Fall webworm).
 OC Neokaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
 OC Arctiidae; Hyphantria.
 OX NCBI_TaxID=39466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98199915; PubMed=9537761;
 RA Park D.S., Shin S.W., Kim M.G., Park S.S., Lee W.J., Brey P.T.,
 RA Park H.Y.;
 RT "Isolation and characterization of the cDNA encoding the
 RT prophenoloxidase of fall webworm, hyphantria cunea."
 RL Insect Biochem. Mol. Biol. 27:983-992(1997).
 DR EMBL; AF020391; AAC34256.1; -.
 DR HSP; P04253; IOXY.
 DR InterPro; IPR000896; Hemocyanin.
 DR InterPro; IPR005203; hemocyanin_C.
 DR InterPro; IPR005204; hemocyanin_N.
 DR InterPro; IPR002227; Tyrosinase.
 DR Pfam; PF00372; hemocyanin_1.
 DR Pfam; PF03723; hemocyanin_C; 1.
 DR Pfam; PF03722; hemocyanin_N; 1.
 DR PRINTS; PR00187; HAEMOCYANIN.
 DR PROSITE; PS00209; HEMOCYANIN_1; 1.
 DR PROSITE; PS00210; HEMOCYANIN_2; 1.
 DR PROSITE; PS00498; TYROSINASE_2; 1.
 SQ SEQUENCE 697 AA; 80641 MW; 86A964E4163D956C CRC64;

Query Match 70.2%; Score 40; DB 5; Length 697;
 Best Local Similarity 44.4%; Pred. No. 1.2e+02;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLYKWHGFV 9
 | | | | : :
 Db 403 FFYRWHAYI 411

RESULT 15
 Q8R946 Q8R946 PRELIMINARY; PRT; 268 AA.
 AC Q8R946;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Pseudouridylylate synthase, 23S RNA-specific.
 DR RLJ2 OR TTE1780.
 OS Thermoanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
 OC Thermoanaerobacteriaceae; Thermoanaerobacter.
 OX NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MB4 / JCM 11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome."
 RL Genome Res. 12:689-700(2002).
 DR EMBL; AE013131; AAM24974.1; -.
 DR InterPro; IPR006145; Pseudou synth.
 DR InterPro; IPR006225; Pseud RLUd.
 DR InterPro; IPR006224; PSI_RLU.
 DR Pfam; PF00849; Pseudou synth_2; 1.
 DR ProDom; PD001819; PSI_RLU; 1.
 DR TIGRFAMs; TIGR00005; FluD_subfam; 1.
 DR PROSITE; PS01129; PSI_RLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 268 AA; 31031 MW; PDE3A13D14FDCFEF CRC64;

Query Match 68.4%; Score 39; DB 16; Length 268;
 Best Local Similarity 83.3%; Pred. No. 67;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LYKWHG 7

Db :|||||

13 MYKWHG 18

Search completed: December 16, 2003, 14:17:57

Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:05 ; Search time 33 Seconds
(without alignments)
43.289 Million cell updates/sec

Title: US-09-870-089b-5

Perfect score: 52

Sequence: 1 FLHKVHFVY 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	9	23	AAU74682
2	43	82.7	1043	22	AAU74683
3	41	78.8	9	23	AAU74685
4	40	76.9	9	23	AAU74684
5	39	75.0	979	22	ABBG3728
6	39	75.0	1044	23	ABP64908
7	38	73.1	102	21	AAAB56395
8	37	71.2	79	23	ABP00720
9	36	69.2	25	15	AAAS7000
					N-terminal fragmen

10	36	69.2	37	15	AAAS7007	N-terminal fragmen
11	36	69.2	95	21	AAG12473	Zea mays protein f
12	36	69.2	128	21	AAG35910	Zea mays protein f
13	36	69.2	166	21	AAG35909	Zea mays protein f
14	36	69.2	182	21	AAG35416	Zea mays protein f
15	36	69.2	195	15	AAG57024	Partial sequence o
16	36	69.2	335	22	AAE05193	Human drug metabol
17	36	69.2	353	12	AAAR11317	Histo-blood gp. A
18	36	69.2	353	15	AAAS7011	Human A transferas
19	36	69.2	353	23	AAE29231	Human blood group
20	36	69.2	353	23	AAO21510	Human blood group
21	36	69.2	354	12	AAAR11789	Histo-blood gp. A
22	36	69.2	354	12	AAAR11790	Histo-blood gp. B
23	36	69.2	354	12	AAAR11792	Histo-blood gp. B
24	36	69.2	354	15	AAAS7013	Human A transferas
25	36	69.2	354	15	AAAS7016	Human B transferas
26	36	69.2	354	15	AAAS7020	Human A transferas
27	36	69.2	354	23	AAE29232	Human blood group
28	36	69.2	354	23	AAO21511	Human blood group
29	36	69.2	358	15	AAAS7014	Human A transferas
30	36	69.2	375	15	AAAS7021	Human A transferas
31	35	67.3	7	22	AAAM44218	H11 binding site c
32	35	67.3	7	22	AAAM45423	H11 binding site c
33	35	67.3	15	23	AAO26506	Human topoisomeras
34	35	67.3	46	23	ABUS1459	Helicobacter pylor
35	35	67.3	53	22	AAMB3305	Human immune/haema
36	35	67.3	72	23	ABUS1593	Helicobacter pylor
37	35	67.3	72	23	ABUS2086	Helicobacter pylor
38	35	67.3	84	23	ABUS1112	Helicobacter pylor
39	35	67.3	84	23	ABUS1190	Helicobacter pylor
40	35	67.3	84	23	ABUS1780	Helicobacter pylor
41	35	67.3	85	23	AAO26505	Human topoisomeras
42	35	67.3	88	23	ABUS1873	Helicobacter pylor
43	35	67.3	94	23	ABUS2204	Helicobacter pylor
44	35	67.3	153	23	ABUS1676	Helicobacter pylor
45	35	67.3	163	23	ABUS0727	Helicobacter pylor

ALIGNMENTS

RESULT 1

AAU74682
ID AAU74682 standard; Peptide; 9 AA.

AC AAU74682;

DT 09-APR-2002 (first entry)

XX Human cancer antigen ATP4/CREB-2 based immunogenic ligand #2.

XX Human; cancer antigen; ATP4; CREB-2; vaccine; cytostatic;

XX immunogenic ligand; gene therapy; MHC; major histocompatibility complex;

XX adoptive immunotherapy; cancer; ovarian cancer.

XX Homo sapiens.

OS Synthetic.

XX WO200192306-A2.

XX 06-DEC-2001.

XX 30-MAY-2001; 2001WO-US17454.

XX 31-MAY-2000; 2000US-209388P.

XX 20-DEC-2000; 2000US-257007P.

XX (GENZ) GENZYME CORP.

XX Nicolette CA;

XX WPI; 2002-097764/13.

PT New therapeutic compounds comprising immunogenic ligands, useful for
PT modulating an immune response, particularly for treating ovarian
PT cancer, and as components of anticancer vaccines -
XX
XX
XX Claim 2; Page 55; 68pp; English.
XX
XX The invention relates to compounds comprising an immunogenic ligand
XX whose sequence is based in part on residues 42-50 of human cancer
XX antigen ATP4/CREB-2 (not defined) and the polynucleotides encoding them.
XX Also included are an antibody that specifically recognises and binds the
XX compound, a method for inducing an immune response in a subject by
XX delivering the compound, a method of immunotherapy comprising
XX administering to a subject the antibody, an immune effector cell that has
XX been raised in vitro or in vivo in the presence and at the expense of an
XX antigen presenting cell that presents the immunogenic compound in the
XX context of an MHC (major histocompatibility complex) molecule and a
XX method of adoptive immunotherapy comprising administering the immune
XX effector cell. The compounds are useful for modulating an immune response
XX to the synthetic and naturally occurring compounds. The compounds are
XX especially useful in gene therapy or as components of anti-cancer
XX vaccines. The compounds are useful for treating cancer, particularly
XX ovarian cancer. The compounds are also useful for generating antibodies
XX that specifically recognise and bind to these molecules. These
XX antibodies are further useful for immunotherapy when administered to a
XX subject. The peptides, polypeptides and polynucleotides are useful in
XX diagnostic methods, for the detection and purification of antibodies,
XX or as immunogens for the production of antibodies. The present
XX sequence represents a human cancer antigen ATP4/CREB-2 based immunogenic
XX ligand of the invention.
XX Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
XX encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
XX but have not been cross-referenced or CDS features put in due to the
XX degeneracy of the DNA sequences.
XX
XX Sequence 9 AA;
XX

Query Match 100.0%; Score 52; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FLHKVHFYV 9
||| |||||
Db 1 FLHKVHFYV 9

RESULT 2
AAG92893
ID AAG92893 standard; Protein; 1043 AA.
XX
XX AAG92893;
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 6647.
XX
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX N-PSDB; AAH68112.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 6647; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX Specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX Sequence 1043 AA;
XX

Query Match 82.7%; Score 43; DB 22; Length 1043;
Best Local Similarity 77.8%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 FLHKVHFYV 9
||| |||||
Db 22 FLHSAHFYV 30

RESULT 3
AAU74685
ID AAU74685 standard; Peptide; 9 AA.
XX
XX AAU74685;
XX
XX 09-APR-2002 (first entry)
XX
XX Human cancer antigen ATP4/CREB-2 based immunogenic ligand #5.
XX
XX Human; cancer antigen; ATF4; CREB-2; vaccine; cytostatic;
XX immunogenic ligand; gene therapy; MHC; major histocompatibility complex;
XX adoptive immunotherapy; cancer; ovarian cancer.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX WO2001192306-A2.
XX
XX 06-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US17454.
XX
XX 31-MAY-2000; 2000US-209388P.
XX
XX 20-DEC-2000; 2000US-257007P.
XX
XX (GENZ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2002-097764/13.
XX
XX New therapeutic compounds comprising immunogenic ligands, useful for
XX modulating an immune response, particularly for treating ovarian
XX cancer, and as components of anticancer vaccines -
XX

PS Claim 5; Page 56; 68pp; English.

XX The invention relates to compounds comprising an immunogenic ligand
CC whose sequence is based in part on residues 42-50 of human cancer
CC antigen ATP4/CREB-2 (not defined) and the polynucleotides encoding them.
CC Also included are an antibody that specifically recognises and binds the
CC compound, a method for inducing an immune response in a subject by
CC delivering the compound, a method of immunotherapy comprising
CC administering to a subject the antibody, an immune effector cell that has
CC been raised in vitro or in vivo in the presence and at the expense of an
CC antigen presenting cell that presents the immunogenic compound in the
CC context of an MHC (major histocompatibility complex) molecule and a
CC method of adoptive immunotherapy comprising administering the immune
CC effector cell. The compounds are useful for modulating an immune response
CC to the synthetic and naturally occurring compounds. The compounds are
CC especially useful in gene therapy or as components of anti-cancer
CC vaccines. The compounds are also useful for generating antibodies
CC that specifically recognise and bind to these molecules. These
CC antibodies are further useful for immunotherapy when administered to a
CC subject. The peptides, polypeptides and polynucleotides are useful in
CC diagnostic methods, for the detection and purification of antibodies,
CC or as immunogens for the production of antibodies. The present
CC sequence represents a human cancer antigen ATP4/CREB-2 based immunogenic
CC ligand of the invention.
CC Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
CC encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
CC but have not been cross-referenced or CDS features put in due to the
CC degeneracy of the DNA sequences.

XX Sequence 9 AA;

Query Match 78.8%; Score 41; DB 23; Length 9;

Best Local Similarity 77.8%; Pred. No. 9.3e+05;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLHKVHFVY 9

Db 1 FLHKVHVLV 9

RESULT 4

AAU74684

ID AAU74684 standard; Peptide; 9 AA.

XX AAU74684;

XX 09-APR-2002 (first entry)

DE Human cancer antigen ATP4/CREB-2 based immunogenic ligand #4.

XX Human; cancer antigen; ATP4; CREB-2; vaccine; cytostatic;
KW immunogenic ligand; gene therapy; MHC; major histocompatibility complex;
KW adoptive immunotherapy; cancer; ovarian cancer.

XX Homo sapiens.

OS Synthetic.

XX WO200192306-A2.

XX 06-DEC-2001.

XX 30-MAY-2001; 2001WO-US17454.

XX 31-MAY-2000; 2000US-209388P.

PR 20-DEC-2000; 2000US-257007P.

XX (GENZ) GENZYME CORP.

XX Nicolette Ca;

XX WPI; 2002-097764/13.

XX

PT New therapeutic compounds comprising immunogenic ligands, useful for
PT modulating an immune response, particularly for treating ovarian
PT cancer, and as components of anticancer vaccines -

PS Claim 4; Page 56; 68pp; English.

XX The invention relates to compounds comprising an immunogenic ligand
CC whose sequence is based in part on residues 42-50 of human cancer
CC antigen ATP4/CREB-2 (not defined) and the polynucleotides encoding them.
CC Also included are an antibody that specifically recognises and binds the
CC compound, a method for inducing an immune response in a subject by
CC delivering the compound, a method of immunotherapy comprising
CC administering to a subject the antibody, an immune effector cell that has
CC been raised in vitro or in vivo in the presence and at the expense of an
CC antigen presenting cell that presents the immunogenic compound in the
CC context of an MHC (major histocompatibility complex) molecule and a
CC method of adoptive immunotherapy comprising administering the immune
CC effector cell. The compounds are useful for modulating an immune response
CC to the synthetic and naturally occurring compounds. The compounds are
CC especially useful in gene therapy or as components of anti-cancer
CC vaccines. The compounds are useful for treating cancer, particularly
CC ovarian cancer. The compounds are also useful for generating antibodies
CC that specifically recognise and bind to these molecules. These
CC antibodies are further useful for immunotherapy when administered to a
CC subject. The peptides, polypeptides and polynucleotides are useful in
CC diagnostic methods, for the detection and purification of antibodies,
CC or as immunogens for the production of antibodies. The present
CC sequence represents a human cancer antigen ATP4/CREB-2 based immunogenic
CC ligand of the invention.

CC Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
CC encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
CC but have not been cross-referenced or CDS features put in due to the
CC degeneracy of the DNA sequences.

XX Sequence 9 AA;

Query Match 76.9%; Score 40; DB 23; Length 9;

Best Local Similarity 77.8%; Pred. No. 9.3e+05;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLHKVHFVY 9

Db 1 FLHKWHVYV 9

RESULT 5

ABB63728

ID ABB63728 standard; Protein; 979 AA.

XX ABB63728;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 17976.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX

DR WPI; 2001-656860/75.
 DR N-PSDB; ABL07831.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 17976; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 979 AA;
 Query Match 75.0%; Score 39; DB 22; Length 979;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FLHKVHF 7
 DQ 663 FLHKVHF 669
 ID ABB64908 standard; Protein; 1044 AA.
 AC ABB64908;
 XX
 XX 25-FEB-2003 (first entry)
 XX
 XX Human protein SEQ ID 568.
 XX
 KW Human; expressed sequence tag; EST;
 KW haematopoietic disorder; central nervous system disease; viral infection;
 KW peripheral nervous system disease; non-healing wound; infectious disease;
 KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
 KW fungal infection; autoimmune disorder; coagulation disorder; noctropic;
 KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
 KW cytostatic; haemostatic; virucide; antibacterial; fungicide;
 KW immunostimulant; cerebroprotective.
 OS Homo sapiens.
 XX
 XX WO200259260-A2.
 XX
 XX 01-AUG-2002.
 XX
 XX 16-NOV-2001; 2001WO-US42950.
 XX
 XX 17-NOV-2000; 2000US-0714936.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
 XX Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX
 XX WPI; 2002-590824/63.
 XX
 XX N-PSDB; ABQ99494.
 XX
 XX New isolated polynucleotide, useful in research, diagnostic or
 PT therapeutic methods, e.g. preventing or treating disorders involving
 PT aberrant protein expression or biological activity -

PS Claim 20; SEQ ID 568; 394pp; English.
 XX
 CC The present invention relates to novel human coding sequences
 CC (ABQ99268-ABQ99608) and proteins (ABP64682-ABP65022). The sequences are
 CC useful in therapeutic, diagnostic and research methods. The
 CC polynucleotides may be used in the field of molecular biology as
 CC hybridisation probes, primers for PCR, for chromosome and gene mapping,
 CC for the recombinant production of protein, or in generation of anti-sense
 CC DNA or RNA. The polynucleotides are useful in diagnostics as expressed
 CC sequence tags (ESTs) for identifying expressed genes or for physical
 CC mapping of the human genome. The proteins may be used as molecular weight
 CC markers, or as nutritional sources or supplements. The proteins may be
 CC used to maintain and expand cell population in a totipotent or
 CC pluripotent state useful for re-engineering damaged or diseased
 CC tissues, transplantation, manufacture of bio-pharmaceuticals or the
 CC development of bio-sensors. The polynucleotides and proteins are useful
 CC for preventing, treating or ameliorating disorders involving aberrant
 CC protein expression or biological activity, e.g. haematopoietic disorders,
 CC central/peripheral nervous system diseases, mechanical and traumatic
 CC disorders, non-healing wounds, immune deficiencies and disorders,
 CC infectious diseases caused by viral, bacterial or fungal infection,
 CC autoimmune disorders, allergic reactions and conditions, coagulation
 CC disorders, or cancer. The polynucleotide sequences of the invention were
 CC assembled from ESTs isolated mainly by sequencing by hybridisation, and
 CC in some cases, sequences obtained from one or more public databases.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1044 AA;
 Query Match 75.0%; Score 39; DB 23; Length 1044;
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHF 8
 DQ 903 FLKHAHFH 910
 ID AAB56395 standard; Protein; 102 AA.
 AC AAB56395;
 XX
 XX 13-MAR-2001 (first entry)
 XX
 XX Human prostate cancer antigen protein sequence SEQ ID NO:973.
 DE
 XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnery; gastrointestinal; nephrotropic; antiinfective; gynaecological;
 KW antibacterial; gene therapy; immune; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease.
 XX
 XX Homo sapiens.
 OS
 XX WO2000055174-A1.
 XX
 XX 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000WO-US05988.
 XX
 XX 12-MAR-1999; 99US-0124270.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX
 XX Rosen CA, Ruben SM;
 XX
 XX WPI; 2000-587513/55.
 DR

DR N-PSDB; AAF15598.
 XX Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX
 XX Claim 11; Page 1413; 2338pp; English.
 XX AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotoxic, antineoplastic, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention.
 XX
 XX Sequence 102 AA;
 SQ
 Query Match 73.1%; Score 38; DB 21; Length 102;
 Best Local Similarity 75.0%; Pred. No. 19;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHFY 8
 Db 4 FEKLHFY 11
 RESULT 8
 ABP00720
 ID ABP00720 standard; Protein; 79 AA.
 XX
 AC ABP00720;
 XX
 XX 24-JUN-2002 (first entry)
 DE Human ORFX protein sequence SEQ ID NO:1422.
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.
 XX
 OS Homo sapiens.
 XX
 PN WO200192523-A2.
 XX
 XX 06-DEC-2001.
 XX
 XX 29-MAY-2001; 2001WO-US10836.
 XX
 XX 30-MAY-2000; 2000US-206132P.
 XX
 XX 29-AUG-2000; 2000US-228716P.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Shinkets RA, Leach MD;
 XX
 XX WPI; 2002-106308/14.
 XX
 XX N-PSDB; AAB16472.
 XX
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT

PT hyperproliferative disorders and autoimmune disorders -
 XX
 XX Disclosure; SEQ ID 1422; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). AAB15762 to AAB27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 79 AA;
 SQ
 Query Match 71.2%; Score 37; DB 23; Length 79;
 Best Local Similarity 44.4%; Pred. No. 23;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 FLHKVHFY 9
 Db 46 FIHSIHYY 54
 RESULT 9
 AAR57000
 ID AAR57000 standard; peptide; 25 AA.
 XX
 AC AAR57000;
 XX
 XX 25-MAR-2003 (updated)
 DT 15-FEB-1995 (first entry)
 XX
 DE N-terminal fragment of human histo-blood group A transferase.
 XX
 KW Blood; group; determinant; antigen; erythrocyte; oligosaccharide;
 KW glycoconjugate; glycosphingolipid; glycoprotein; glycosyltransferase;
 KW transferase.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 15 /label= Unsure.
 FT /note= "This amino acid is unknown."
 FT Misc-difference 16 /label= Unsure.
 FT /note= "This amino acid is unknown."
 FT
 XX US5326857-A.
 XX
 XX 05-JUL-1994.
 PD
 XX 29-AUG-1991; 91US-0752101.
 XX
 XX 31-AUG-1989; 89US-0402695.
 PR

```

PR 29-AUG-1991; 91US-0752101.
XX (BIOM-) BIOMEMBRANE INST.
XX Clausen H, Hakomori S, White T, Yamamoto F;
XX WPI; 1994-217098/26.
XX Isolated DNA molecules - encode human histo-blood groups A-, B-
XX and O-glycotransferases
XX Example 2; Column 35-36; 63pp; English.
XX The histo-blood group ABH determinants are major allogeneic antigens
XX in both erythrocytes and tissues of humans. They generally
XX constitute peripheral parts of the oligosaccharide chains of
XX glycoconjugates i.e. linked to lipids (glycosphingolipids) or to
XX proteins (glycoproteins). It was proposed that the A and B
XX phenotypes were associated with glycosyltransferases that converted
XX the H substance associated with the O phenotype to A and B
XX respectively, through the addition of alpha1-3-N-acetylgalactosamine
XX or alpha1-3-galactosyl residues to the H antigen Fuc-alpha1-2Gal-
XX beta1-R. Hence, the primary products of the histo-blood group A
XX and B genes are the respective glycosyltransferases. This is a
XX fragment of the A group transferase. See also AAR56995-R57010.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX Sequence 25 AA;
SQ Query Match 69.2%; Score 36; DB 15; Length 25;
Best Local Similarity 71.4%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFVY 9
Db .:|:|:|
6 HRVHYV 12

RESULT 10
AAR57007
ID AAR57007 standard; peptide; 37 AA.
XX
XX AAR57007;
AC AAR57007;
XX
XX 25-MAR-2003 (updated)
DT 15-FEB-1995 (first entry)
XX
XX N-terminal fragment of human histo-blood group A transferase.
XX
XX Blood; group; determinant; antigen; erythrocyte; oligosaccharide;
XX glycoconjugate; glycosphingolipid; glycoprotein; glycosyltransferase;
XX transferase.
XX
XX Homo sapiens.
XX US5326857-A.
XX
XX 05-JUL-1994.
XX
XX 29-AUG-1991; 91US-0752101.
XX
XX 31-AUG-1989; 89US-0402695.
XX
XX 29-AUG-1991; 91US-0752101.
XX
XX (BIOM-) BIOMEMBRANE INST.
XX
XX Clausen H, Hakomori S, White T, Yamamoto F;
XX WPI; 1994-217098/26.
XX
XX Isolated DNA molecules - encode human histo-blood groups A-, B-
XX and O-glycotransferases
XX Example 2; Column 35-36; 63pp; English.
XX The histo-blood group ABH determinants are major allogeneic antigens
XX in both erythrocytes and tissues of humans. They generally
XX constitute peripheral parts of the oligosaccharide chains of
XX glycoconjugates i.e. linked to lipids (glycosphingolipids) or to
XX proteins (glycoproteins). It was proposed that the A and B
XX phenotypes were associated with glycosyltransferases that converted
XX the H substance associated with the O phenotype to A and B
XX respectively, through the addition of alpha1-3-N-acetylgalactosamine
XX or alpha1-3-galactosyl residues to the H antigen Fuc-alpha1-2Gal-
XX beta1-R. Hence, the primary products of the histo-blood group A
XX and B genes are the respective glycosyltransferases. This is a
XX fragment of the A group transferase. See also AAR56995-R57010.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX Sequence 37 AA;
SQ Query Match 69.2%; Score 36; DB 15; Length 37;
Best Local Similarity 71.4%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFVY 9
Db .:|:|:|
3 HRVHYV 9

RESULT 11
AAG12473
ID AAG12473 standard; Protein; 95 AA.
XX
XX AAG12473;
AC AAG12473;
XX
XX 17-OCT-2000 (first entry)
DT
XX
XX Zea mays protein fragment SEQ ID NO: 11599.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence; corn.
XX
XX Zea mays subsp. mays.
XX
XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 03-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126264.
XX
XX 29-MAR-1999; 99US-0126785.
XX
XX 01-APR-1999; 99US-0127462.
XX
XX 06-APR-1999; 99US-0128234.
XX
XX 08-APR-1999; 99US-0128714.
XX
XX 16-APR-1999; 99US-0129845.
XX
XX 19-APR-1999; 99US-0130077.
XX
XX 21-APR-1999; 99US-0130449.
XX
XX 23-APR-1999; 99US-0130510.
XX
XX 28-APR-1999; 99US-0130891.
XX
XX 30-APR-1999; 99US-0131449.
XX
XX 30-APR-1999; 99US-0132048.
XX
XX 30-APR-1999; 99US-0132407.
XX
XX 04-MAY-1999; 99US-0132484.
XX
XX 05-MAY-1999; 99US-0132485.
XX
XX 06-MAY-1999; 99US-0132486.
XX
XX 06-MAY-1999; 99US-0132487.
XX
XX 07-MAY-1999; 99US-0132863.
XX
XX 11-MAY-1999; 99US-0134256.
XX
XX 14-MAY-1999; 99US-0134218.

```

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160999.

```

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161923.
PR 28-OCT-1999; 99US-0161924.
PR 29-OCT-1999; 99US-0162142.

Query Match 69.2%; Score 36; DB 21; Length 95;
Best Local Similarity 55.6%; Pred. No. 41;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLHKVHFV 9
Db 74 FVHDLHFFV 82

RESULT 12
AAG25910
ID AAG25910 standard; Protein; 128 AA.
XX AC AAG25910;
XX DT 17-OCT-2000 (first entry)
XX DE Zea mays protein fragment SEQ ID NO: 30165.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

```



```
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 69.3%; Score 36; DB 21; Length 128;
Best Local Similarity 55.6%; Pred. No. 55;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFV 9
Db 74 FVHDLHFFV 82

RESULT 13
AAG25909
ID AAG25909 standard; Protein; 166 AA.
AC AAG25909;
XX
XX 17-OCT-2000 (first entry)
DE Zea mays protein fragment SEQ ID NO: 30164.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
```

```
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145091.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 69.2%; Score 36; DB 21; Length 166;
Best Local Similarity 55.6%; Pred.No. 72;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 FLKHVHFV 9
Db 112 FVHDLHFV 120

RESULT 14
AAG35416
ID AAG35416 standard; Protein; 182 AA.
XX AC AAG35416;
XX DT 18-OCT-2000 (first entry)
XX DE Zea mays protein fragment SEQ ID NO: 43259.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147301.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0149684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162143.

Query Match 69.2%; Score 36; DB 21; Length 182;
 Best Local Similarity 55.6%; Pred. No. 79;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLKHVHYV 9
 | : | : | : |
 Db 74 FVHDLHFFV 82

RESULT 15

AAR57024
 ID AAR57024 standard; Protein; 195 AA.
 XX AC AAR57024;
 XX 25-MAR-2003 (updated)
 DT 16-FEB-1995 (first entry)
 XX Partial sequence of human A transferase.
 XX Blood; group; determinant; antigen; erythrocyte; oligosaccharide;
 KW glycoconjugate; glycosphingolipid; glycoprotein; glycosyltransferase;
 transferase.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT Region 3
 FT /note= "Site of deletion."
 XX US5326857-A.
 PD 05-JUL-1994.
 XX 29-AUG-1991; 91US-0752101.
 XX 31-AUG-1989; 89US-0402695.
 XX 29-AUG-1991; 91US-0752101.
 XX (BIOM-) BIOMEMBRANE INST.
 XX Clausen H, Hakomori S, White T, Yamamoto F;
 WPI; 1994-217098/26.
 XX N-PSDB; AAQ68826.
 XX Isolated DNA molecules - encode human histo-blood groups A-, B-
 and O-glycotransferases
 XX Example 9; Figure 10; 63pp; English.
 XX The histo-blood group ABH determinants are major allogeneic antigens
 in both erythrocytes and tissues of humans. They generally
 constitute peripheral parts of the oligosaccharide chains of
 glycoconjugates i.e. linked to lipids (glycosphingolipids) or to
 proteins (glycoproteins). It was proposed that the A and B
 phenotypes were associated with glycosyltransferases that converted
 the H substance associated with the O phenotype to A and B
 respectively, through the addition of alpha1-3-N-acetylgalactosamine
 or alpha1-3-galactosyl residues to the H antigen Fuc-alpha1-2gal-
 beta1-K. Hence, the primary products of the histo-blood group A
 and B genes are the respective glycosyltransferases. The full
 sequence of the human transferase A is described in AAR57011.
 XX (Updated on 25-MAR-2003 to correct PF field.)

SQ Sequence 195 AA;

Query Match 69.2%; Score 36; DB 15; Length 195;
 Best Local Similarity 71.4%; Pred. No. 84;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHYV 9
 | : | : | : |
 Db 20 HRVHYV 26

Search completed: December 16, 2003, 14:14:27
 Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:11:00 ; Search time 12.3333 Seconds
(without alignments)
30.875 Million cell updates/sec

Title: US-09-870-089B-5
Sequence: 1 FLHKVHFVY 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	69.2	25	1	US-07-752-101A-6
2	36	69.2	37	1	US-07-752-101A-13
3	36	69.2	195	1	US-07-752-101A-68
4	36	69.2	353	1	US-07-752-101A-35
5	36	69.2	353	1	US-07-752-101A-36
6	36	69.2	354	1	US-07-752-101A-38
7	36	69.2	354	1	US-07-752-101A-39
8	36	69.2	354	1	US-07-752-101A-41
9	36	69.2	354	1	US-07-752-101A-51
10	36	69.2	375	1	US-07-752-101A-52
11	35	67.3	372	1	US-08-202-056-5
12	35	67.3	372	1	US-08-076-093A-6
13	35	67.3	372	1	US-08-701-245-6
14	35	67.3	372	2	US-08-284-586-6
15	35	67.3	372	2	US-08-805-478-6
16	35	67.3	372	2	US-08-802-627A-6
17	35	67.3	372	2	US-08-801-238-6
18	35	67.3	372	2	US-08-801-238-6
19	35	67.3	372	3	US-09-104-296-6
20	35	67.3	372	3	US-08-982-433-8
21	35	67.3	372	4	US-09-170-496D-66
22	35	67.3	372	4	US-09-170-496D-200
23	35	67.3	705	4	US-09-328-352-7436
24	34	65.4	105	2	US-08-889-013C-6
25	34	65.4	311	1	US-08-872-784-1
26	34	65.4	311	2	US-09-100-851-1
27	34	65.4	311	3	US-09-265-294-1

Sequence 6, Appli
Sequence 18, Appl
Sequence 5000, Ap
Sequence 4, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 4499, Ap
Sequence 5382, Ap
Sequence 14, Appl
Sequence 24497, A
Sequence 313, App
Sequence 7, Appli
Sequence 313, App
Sequence 48, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 671, App
Sequence 4203, Ap

ALIGNMENTS

RESULT 1
US-07-752-101A-6
; Sequence 6, Application US/07752101A
; Patent No. 5326857
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Fumi-ichiro
; APPLICANT: White, Thayer
; APPLICANT: Hakomori, Sen-itiroh
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: ABO GENOTYPING
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07752,101A
; FILING DATE: 19910829
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 150036.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: Region
; LOCATION: 15..16
; OTHER INFORMATION: /label= unsure
; OTHER INFORMATION: /note= "These amino acids are unknown."

US-07-752-101A-6

Query Match 69.2%; Score 36; DB 1; Length 25;
 Best Local Similarity 71.4%; Pred. No. 4.9;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
 ||:|:|
 Db 6 HRVHYV 12

RESULT 2

US-07-752-101A-13
 ; Sequence 13, Application US/07752101A
 ; Patent No. 5326857
 ; GENERAL INFORMATION:
 ; APPLICANT: Yamamoto, Fumi-ichiro
 ; APPLICANT: White, Thayer
 ; APPLICANT: Hakomori, Sen-itiroh
 ; APPLICANT: Clausen, Henrik
 ; TITLE OF INVENTION: ABO GENOTYPING
 ; NUMBER OF SEQUENCES: 69
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.
 ; ZIP: 98104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07752,101A
 ; FILING DATE: 19910829
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sharkey, Richard G.
 ; REGISTRATION NUMBER: 32,629
 ; REFERENCE/DOCKET NUMBER: 150036.406C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-622-4900
 ; TELEFAX: 206-682-6031
 ; TELEX: 3723836
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 37 amino acids
 ; TYPE: AMINO ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; US-07-752-101A-13

Query Match 69.2%; Score 36; DB 1; Length 37;
 Best Local Similarity 71.4%; Pred. No. 7.1;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
 ||:|:|
 Db 3 HRVHYV 9

RESULT 3

US-07-752-101A-68
 ; Sequence 68, Application US/07752101A
 ; Patent No. 5326857
 ; GENERAL INFORMATION:
 ; APPLICANT: Yamamoto, Fumi-ichiro

; APPLICANT: White, Thayer
 ; APPLICANT: Hakomori, Sen-itiroh
 ; APPLICANT: Clausen, Henrik
 ; TITLE OF INVENTION: ABO GENOTYPING
 ; NUMBER OF SEQUENCES: 69
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.
 ; ZIP: 98104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07752,101A
 ; FILING DATE: 19910829
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sharkey, Richard G.
 ; REGISTRATION NUMBER: 32,629
 ; REFERENCE/DOCKET NUMBER: 150036.406C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-622-4900
 ; TELEFAX: 206-682-6031
 ; TELEX: 3723836
 ; INFORMATION FOR SEQ ID NO: 68:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 195 amino acids
 ; TYPE: AMINO ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; US-07-752-101A-68

Query Match 69.2%; Score 36; DB 1; Length 195;
 Best Local Similarity 71.4%; Pred. No. 34;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
 ||:|:|
 Db 20 HRVHYV 26

RESULT 4

US-07-752-101A-35
 ; Sequence 35, Application US/07752101A
 ; Patent No. 5326857
 ; GENERAL INFORMATION:
 ; APPLICANT: Yamamoto, Fumi-ichiro
 ; APPLICANT: White, Thayer
 ; APPLICANT: Hakomori, Sen-itiroh
 ; APPLICANT: Clausen, Henrik
 ; TITLE OF INVENTION: ABO GENOTYPING
 ; NUMBER OF SEQUENCES: 69
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.
 ; ZIP: 98104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/752,101A
; FILING DATE: 19910829
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 150036.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 353 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-752-101A-35
Query Match 69.2%; Score 36; DB 1; Length 353;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 3 HKVHFYV 9
Db 144 HRVHYV 150

```

```

RESULT 5
US-07-752-101A-36
; Sequence 36, Application US/07752101A
; Patent No. 5326857
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Fumi-ichiro
; APPLICANT: White, Thayer
; APPLICANT: Hakomori, Sen-itiroh
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: ABO GENOTYPING
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/752,101A
; FILING DATE: 19910829
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 150036.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 353 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

```

; FRAGMENT TYPE: N-terminal
; US-07-752-101A-36
Query Match 69.2%; Score 36; DB 1; Length 353;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 3 HKVHFYV 9
Db 144 HRVHYV 150

```

```

RESULT 6
US-07-752-101A-38
; Sequence 38, Application US/07752101A
; Patent No. 5326857
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Fumi-ichiro
; APPLICANT: White, Thayer
; APPLICANT: Hakomori, Sen-itiroh
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: ABO GENOTYPING
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/752,101A
; FILING DATE: 19910829
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 150036.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-07-752-101A-38
Query Match 69.2%; Score 36; DB 1; Length 354;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 3 HKVHFYV 9
Db 145 HRVHYV 151

```

```

RESULT 7
US-07-752-101A-39
; Sequence 39, Application US/07752101A
; Patent No. 5326857
; GENERAL INFORMATION:

```

```

; APPLICANT: Yamamoto, Fumi-ichiro
; APPLICANT: White, Thayer
; APPLICANT: Hakomori, Sen-itiroh
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: ABO GENOTYPING
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/752,101A
; FILING DATE: 19910829
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 150036.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-07-752-101A-39

```

```

Query Match          69.2%; Score 36; DB 1; Length 354;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 HKVHFYV 9
       |:|:|:|
Db      145 HRVHYV 151

```

```

RESULT 8
US-07-752-101A-41
; Sequence 41, Application US/07752101A
; Patent No. 5326857
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Fumi-ichiro
; APPLICANT: White, Thayer
; APPLICANT: Hakomori, Sen-itiroh
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: ABO GENOTYPING
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/752,101A
; FILING DATE: 19910829
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 150036.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-07-752-101A-41

```

```

Query Match          69.2%; Score 36; DB 1; Length 354;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 HKVHFYV 9
       |:|:|:|
Db      145 HRVHYV 151

```

```

RESULT 9
US-07-752-101A-51
; Sequence 51, Application US/07752101A
; Patent No. 5326857
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Fumi-ichiro
; APPLICANT: White, Thayer
; APPLICANT: Hakomori, Sen-itiroh
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: ABO GENOTYPING
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/752,101A
; FILING DATE: 19910829
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 150036.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: AMINO ACID

```


STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 US-07-752-101A-51

Query Match 69.2%; Score 36; DB 1; Length 354;
 Best Local Similarity 71.4%; Pred. No. 60;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HKVHFYV 9
 Db 145 HRVHYV 151

RESULT 10

US-07-752-101A-52
 Sequence 52, Application US/07752101A
 Patent No. 5326857
 GENERAL INFORMATION:
 APPLICANT: Yamamoto, Fumi-ichiro
 APPLICANT: White, Thayer
 APPLICANT: Hakomori, Sen-itiroh
 APPLICANT: Clausen, Henrik
 TITLE OF INVENTION: ABO GENOTYPING
 NUMBER OF SEQUENCES: 69
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berty
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.
 ZIP: 98104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/752,101A
 FILING DATE: 19910829
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Sharkey, Richard G.
 REGISTRATION NUMBER: 32,629
 REFERENCE/DOCKET NUMBER: 150036.406C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 206-622-4900
 TELEFAX: 206-692-6031
 TELEX: 3723836
 INFORMATION FOR SEQ ID NO: 52:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 375 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 FEATURE:
 NAME/KEY: Region
 LOCATION: 1..53
 OTHER INFORMATION: /label= unsure
 OTHER INFORMATION: /note= "these amino acids are unknown."
 US-07-752-101A-52

Query Match 69.2%; Score 36; DB 1; Length 375;
 Best Local Similarity 71.4%; Pred. No. 64;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HKVHFYV 9
 Db 145 HRVHYV 151

RESULT 11

US-08-202-056-5
 Sequence 5, Application US/08202056
 Patent No. 5440021
 GENERAL INFORMATION:
 APPLICANT: Chuntharapai, Anan
 APPLICANT: Hebert, Caroline
 APPLICANT: Kim, Kyung Jin
 APPLICANT: Lee, James
 TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: patin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/202,056
 FILING DATE: 25-FEB-1994
 CLASSIFICATION: 436
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/677211
 FILING DATE: 29-MAR-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Love, Richard B.
 REGISTRATION NUMBER: 34,659
 REFERENCE/DOCKET NUMBER: 706P3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-5530
 TELEFAX: 415/952-9881
 TELEX: 910/371-7168
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 372 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-202-056-5

Query Match 67.1%; Score 35; DB 1; Length 372;
 Best Local Similarity 85.7%; Pred. No. 94;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKVHFY 8
 Db 128 LHKVNFY 134

RESULT 12

US-08-076-093A-6
 Sequence 6, Application US/08076093A
 Patent No. 5543503
 GENERAL INFORMATION:
 APPLICANT: Chuntharapai, Anan
 APPLICANT: Lee, James
 APPLICANT: Hebert, Caroline
 APPLICANT: Jin Kim, K.
 TITLE OF INVENTION: Antibodies to Human PF4A Receptors
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd

```

; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,093A
; FILING DATE: 11-Jun-1993
; CLASSIFICATION: 530
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
;
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
; US-08-076-093A-6

```

```

Query Match 67.3%; Score 35; DB 1; Length 372;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 LHKVHFY 8
Db 128 LHKVNFY 134

```

```

RESULT 13
US-08-701-265-6
; Sequence 6, Application US/08701265
; Patent No. 5776457
;
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
;
; TITLE OF INVENTION: Antibodies to Human PF4A Receptors
;
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/701,265
; FILING DATE: 22-AUG-1996
; CLASSIFICATION: 424
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076093
; FILING DATE: 11-Jun-1993

```

```

; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
; US-08-701-265-6

```

```

Query Match 67.3%; Score 35; DB 1; Length 372;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 LHKVHFY 8
Db 128 LHKVNFY 134

```

```

RESULT 14
US-08-284-586-6
; Sequence 6, Application US/08284586
; Patent No. 5840856
;
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
;
; TITLE OF INVENTION: Antibodies to Human PF4A Receptors
;
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,586
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,093A
; FILING DATE: 11-Jun-1993
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
;
; INFORMATION FOR SEQ ID NO: 6:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-284-586-6

Query Match 67.3%; Score 35; DB 2; Length 372;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHFY 8
||||:|
Db 128 LHKVNFY 134

RESULT 15
US-08-805-478-6
Sequence 6, Application US/08805478
Patent No. 5874543
GENERAL INFORMATION:
APPLICANT: Chuntharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: ANTIBODIES TO PF4A RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,478
FILING DATE: 25-Feb-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-805-478-6

Query Match 67.3%; Score 35; DB 2; Length 372;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHFY 8
||||:|
Db 128 LHKVNFY 134

Search completed: December 16, 2003, 14:20:39
Job time : 12.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:18:11 ; Search time 22.1667 Seconds
(without alignments)
75.512 Million cell updates/sec

Title: US-09-870-089B-5

Perfect score: 52

Sequence: 1 FLHKVHFYV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	100.0	9	12	US-09-870-089B-5
2	43	82.7	1043	10	US-09-738-626-6647
3	41	78.8	9	12	US-09-870-089B-11
4	40	76.9	9	12	US-09-870-089B-9
5	39	75.0	942	12	US-10-015-115-107
6	38	73.1	102	10	US-09-925-300-973
7	36	69.2	353	11	US-09-994-427A-3
8	36	69.2	353	11	US-09-995-419A-11
9	36	69.2	353	15	US-10-105-963-12
10	36	69.2	354	10	US-09-994-427A-5
11	36	69.2	354	11	US-09-995-419A-13
12	36	69.2	354	15	US-10-105-963-14
13	35	67.3	371	9	US-09-815-242-11311
14	35	67.3	371	9	US-09-815-242-11490
15	35	67.3	371	10	US-09-895-913A-298

35	67.3	372	10	US-09-104-063-6	Sequence 6, Appli
35	67.3	372	12	US-10-239-423-80	Sequence 80, Appli
35	67.3	372	15	US-10-251-385-66	Sequence 66, Appli
35	67.3	372	15	US-10-251-385-200	Sequence 200, App
35	67.3	372	15	US-10-225-567A-60	Sequence 60, Appli
34	65.4	113	11	US-09-764-891-4446	Sequence 4446, Ap
34	65.4	178	9	US-09-815-242-13494	Sequence 13494, A
34	65.4	311	9	US-09-766-366-1	Sequence 1, Appli
34	65.4	428	15	US-10-157-031-139	Sequence 139, App
34	65.4	1802	10	US-09-965-553-18	Sequence 18, Appli
33	63.5	58	15	US-10-157-031-273	Sequence 273, App
33	63.5	106	9	US-09-925-302-640	Sequence 640, Appli
33	63.5	189	14	US-10-046-643-6	Sequence 6, Appli
33	63.5	366	9	US-09-819-946-6	Sequence 10, Appli
33	63.5	626	9	US-09-801-574-10	Sequence 4, Appli
33	63.5	763	9	US-09-819-946-4	Sequence 3, Appli
33	63.5	777	11	US-09-361-652-3	Sequence 3, Appli
33	63.5	777	11	US-09-927-315-3	Sequence 3, Appli
33	63.5	777	15	US-10-225-567A-676	Sequence 676, App
33	63.5	840	12	US-10-246-785-1	Sequence 1, Appli
33	63.5	840	12	US-10-190-417-3	Sequence 3, Appli
33	63.5	841	9	US-09-819-946-2	Sequence 2, Appli
33	63.5	841	10	US-09-897-427A-2	Sequence 2, Appli
33	63.5	841	11	US-09-799-629-17	Sequence 17, Appli
33	63.5	841	12	US-10-130-417-27	Sequence 27, Appli
33	63.5	841	15	US-10-035-045-17	Sequence 17, Appli
32	61.5	9	12	US-09-870-089B-7	Sequence 7, Appli
32	61.5	46	11	US-09-764-891-4433	Sequence 4433, Ap
32	61.5	56	15	US-10-007-280A-172	Sequence 172, App
32	61.5	78	9	US-09-764-869-1039	Sequence 1039, Ap

ALIGNMENTS

RESULT 1

US-09-870-089B-5
; Sequence 5, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870, 089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATP4/CREB-2
US-09-870-089B-5

Query Match 100.0%; Score 52; DB 12; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHFYV 9

DB 1 FLHKVHFYV 9

RESULT 2

US-09-738-626-6647
; Sequence 6647, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO

```

; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6647
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6647

Query Match      82.7%; Score 43; DB 10; Length 1043;
Best Local Similarity 77.8%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 FLHKVHFVY 9
      |||||
DB      22 FLHSAHFVY 30

RESULT 3
US-09-870-089B-11
; Sequence 11, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATF4/CREB-2
US-09-870-089B-11

Query Match      78.8%; Score 41; DB 12; Length 9;
Best Local Similarity 77.8%; Pred. No. 6.1e+05;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 FLHKVHFVY 9
      |||||
DB      1 FLHKVHFVY 9

RESULT 4
US-09-870-089B-9
; Sequence 9, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATF4/CREB-2
US-09-870-089B-9

Query Match      75.0%; Score 39; DB 12; Length 942;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FLHKVHF 7
      |||||
DB      622 FLHKVHF 628

US-10-015-115-107
; Sequence 107, Application US/10015115
; Publication No. US20030207800A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesha
; APPLICANT: Gangolli, Esha A
; APPLICANT: Shimkets, Richard A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Mutalichara
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-211
; CURRENT APPLICATION NUMBER: US/10/015,115
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: 60/248,153
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/249,598
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/264,240
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/266,127
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/269,562
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/304,348
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/309,261
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,283
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 107
; LENGTH: 942
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-015-115-107

Query Match      75.0%; Score 39; DB 12; Length 942;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FLHKVHF 7
      |||||
DB      622 FLHKVHF 628

```

RESULT 6

US-09-925-300-973
 ; Sequence 973, Application US/09925300
 ; Patent No. US20020151681A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Craig Rosen,
 ; APPLICANT: Steve Ruben,
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA101
 ; CURRENT APPLICATION NUMBER: US/09/925,300
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05988
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1890
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 973
 ; LENGTH: 102
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-925-300-973

Query Match 73.1%; Score 38; DB 10; Length 102;
 Best Local Similarity 75.0%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLKHVHFY 8
 Db 4 FEHKLHFY 11

RESULT 7

US-09-994-427A-3
 ; Sequence 3, Application US/09994427A
 ; Patent No. US20020128221A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Geron Corporation
 ; APPLICANT: Schiff, J. Michael
 ; TITLE OF INVENTION: GLYCOSYLTRANSFERASE VECTORS FOR TREATING CANCER
 ; FILE REFERENCE: 083,002
 ; CURRENT APPLICATION NUMBER: US/09/994,427A
 ; CURRENT FILING DATE: 2002-02-26
 ; PRIOR APPLICATION NUMBER: 60/253,395
 ; PRIOR FILING DATE: 2000-11-27
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 353
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-994-427A-3

Query Match 69.2%; Score 36; DB 10; Length 353;
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 3 HKVHFYV 9
 Db 144 HRVHYV 150

RESULT 8

US-09-995-419A-11
 ; Sequence 11, Application US/09995419A
 ; Publication No. US2003032187A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Geron Corporation
 ; APPLICANT: McWhir, Jim
 ; APPLICANT: Gold, Joseph D.
 ; APPLICANT: Schiff, J. Michael
 ; TITLE OF INVENTION: 096,004 - SeqList
 ; FILE REFERENCE: 096,004 - SeqList

; CURRENT APPLICATION NUMBER: US/09/995,419A
 ; CURRENT FILING DATE: 2001-11-26
 ; PRIOR APPLICATION NUMBER: 60/253,357
 ; PRIOR FILING DATE: 2000-11-27
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 11
 ; LENGTH: 353
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-995-419A-11

Query Match 69.2%; Score 36; DB 11; Length 353;
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 3 HKVHFYV 9
 Db 144 HRVHYV 150

RESULT 9

US-10-105-963-12
 ; Sequence 12, Application US/10105963
 ; Publication No. US2003008818A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Geron Corporation
 ; APPLICANT: Denning, Chris
 ; APPLICANT: Clark, A. John
 ; APPLICANT: Schiff, J. Michael
 ; TITLE OF INVENTION: Animal Tissue with Carbohydrate Antigens Compatible for Human
 ; TITLE OF INVENTION: Transplantation and a Carbohydrate Determinant Selection System
 ; FILE REFERENCE: 731/002
 ; CURRENT APPLICATION NUMBER: US/10/105,963
 ; CURRENT FILING DATE: 2002-03-21
 ; PRIOR APPLICATION NUMBER: US 60/277,811
 ; PRIOR FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 12
 ; LENGTH: 353
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-105-963-12

Query Match 69.2%; Score 36; DB 15; Length 353;
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 3 HKVHFYV 9
 Db 144 HRVHYV 150

RESULT 10

US-09-994-427A-5
 ; Sequence 5, Application US/09994427A
 ; Patent No. US20020128221A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Geron Corporation
 ; APPLICANT: Schiff, J. Michael
 ; TITLE OF INVENTION: GLYCOSYLTRANSFERASE VECTORS FOR TREATING CANCER
 ; FILE REFERENCE: 083,002
 ; CURRENT APPLICATION NUMBER: US/09/994,427A
 ; CURRENT FILING DATE: 2002-02-26
 ; PRIOR APPLICATION NUMBER: 60/253,395
 ; PRIOR FILING DATE: 2000-11-27
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 354
 ; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-994-427A-5

Query Match 69.2%; Score 36; DB 10; Length 354;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
|:|:|:
Db 145 HRVHYV 151

RESULT 11
US-09-995-419A-13
; Sequence 13, Application US/09995419A
; Publication No. US20030032187A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: McWhir, Jim
; APPLICANT: Gold, Joseph D.
; APPLICANT: Schiff, J. Michael
; FILE REFERENCE: 096,004 - SeqList
; CURRENT APPLICATION NUMBER: US/09/995,419A
; CURRENT FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 60/253,357
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 13
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-995-419A-13

Query Match 69.2%; Score 36; DB 11; Length 354;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
|:|:|:
Db 145 HRVHYV 151

RESULT 12
US-10-105-963-14
; Sequence 14, Application US/10105963
; Publication No. US20030068818A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Denning, Chris
; APPLICANT: Clark, A. John
; APPLICANT: Schiff, J. Michael
; TITLE OF INVENTION: Animal Tissue with Carbohydrate Antigens Compatible for Human
; TITLE OF INVENTION: Transplantation and a Carbohydrate Determinant Selection System
; FILE REFERENCE: 731/002
; CURRENT APPLICATION NUMBER: US/10/105,963
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 60/277,811
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 14
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-963-14

Query Match 69.2%; Score 36; DB 15; Length 354;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
|:|:|:
Db 145 HRVHYV 151

RESULT 13
US-09-815-242-11311
; Sequence 11311, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11311
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11311

Query Match 67.3%; Score 35; DB 9; Length 371;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHFYV 9
|:|:|:
Db 333 LHNLFYL 340

RESULT 14
US-09-815-242-11490
; Sequence 11490, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 80/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11490
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11490

Query Match 67.3%; Score 35; DB 9; Length 371;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHFV 9
DB 333 LHNLFYL 340

RESULT 15
US-09-895-913A-298
; Sequence 298, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 298
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-298

Query Match 67.3%; Score 35; DB 10; Length 371;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHFV 9
DB 333 LHNLFYL 340

Search completed: December 16, 2003, 14:41:05
Job time : 22.1667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:10:15 ; Search time 10.3333 Seconds
(without alignments)
83.760 Million cell updates/sec

Title: US-09-870-089b-5

Perfect score: 52

Sequence: 1 FLHKVHFYV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 76:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	37	71.2	660	2 T22794	hypothetical prote
2	36	69.2	116	2 S50955	hypothetical prote
3	36	69.2	295	2 PC1120	glycoprotein-fucos
4	36	69.2	353	1 A34933	glycoprotein-fucos
5	36	69.2	363	2 T46126	hypothetical prote
6	36	69.2	412	2 T46104	hypothetical prote
7	36	69.2	512	2 S60622	cytochrome-c oxida
8	35	67.3	250	2 AH0655	probable regulator
9	35	67.3	251	2 T32200	hypothetical prote
10	35	67.3	327	2 S56162	MDCR15 protein - h
11	35	67.3	371	2 C71952	queuine tRNA-ribos
12	35	67.3	371	2 A64555	tRNA-guanine trans
13	35	67.3	372	2 S26867	G protein-coupled
14	35	67.3	426	2 F64419	hypothetical prote
15	35	67.3	569	2 T03390	4-coumarate-CoA li
16	35	67.3	642	2 A82428	sensory box/GGDEF
17	35	67.3	650	2 T00094	endostyle-specific
18	34	65.4	139	2 S41450	self-incompatibili
19	34	65.4	159	2 A90546	hypothetical prote
20	34	65.4	178	2 B95183	non-heme iron-cont
21	34	65.4	178	2 E98050	hypothetical prote
22	34	65.4	304	2 T22597	hypothetical prote
23	34	65.4	314	2 T31866	hypothetical prote
24	34	65.4	319	1 JC5644	acyl-CoA thioleste
25	34	65.4	374	2 S42628	G protein-coupled
26	34	65.4	374	2 S32785	G protein-coupled
27	34	65.4	1346	2 G71613	hypothetical prote
28	34	65.4	1353	2 T19691	hypothetical prote
29	34	65.4	1375	2 T37672	probable DNA repai

30 33 63.5 153 2 G69660 metal-regulated pr
31 33 63.5 190 2 S16817 proteinase 1, mito
32 33 63.5 229 2 T27466 hypothetical prote
33 33 63.5 243 2 B72757 probable phosphate
34 33 63.5 253 2 H81391 amino-acid ABC tra
35 33 63.5 261 2 T05665 hypothetical prote
36 33 63.5 265 2 T15451 hypothetical prote
37 33 63.5 300 2 T38986 probable c-4 methy
38 33 63.5 331 2 A71365 probable Lambda CI
39 33 63.5 343 2 A71365 B. subtilis YkpP p
40 33 63.5 347 2 F84773 B. subtilis YkpP p
41 33 63.5 368 2 AH3202 conserved hypotet
42 33 63.5 430 2 T23899 conserved hypotet
43 33 63.5 432 1 I39877 hypothetical prote
44 33 63.5 589 2 G83758 diaminopimelate de
45 33 63.5 oligopeptidase

ALIGNMENTS

RESULT 1

T22794
hypothetical protein F56H6.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22794

R:Kershaw, J.

submitted to the EMBL Data Library, November 1996
A:Reference number: Z19617

A:Accession: T22794
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-660 <WIL>

A:Cross-references: EMBL:Z81553; PIDN:CAB04490.1; GSPDB:GN00019; CESP:F56H6.1
A:Experimental source: clone F56H6

C:Genetics:

A:Gene: CESP:F56H6.1

A:Map position: 1

A:Introns: 34/2; 54/2; 197/3; 283/3; 323/3; 384/1; 404/3; 452/2; 493/2; 538/2; 591/1

Query Match 71.2%; Score 37; DB 2; Length 660;

Best Local Similarity 62.5%; Pred. No. 41;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKVHFY 8
|||||:
Db 620 FLHKVHF 627

RESULT 2

S50955
hypothetical protein YLL065w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein L0536

C:Species: Saccharomyces cerevisiae

C:Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 20-Jun-2000

C:Accession: S50955; S64817

R:Wedler, H.; Wambutt, R.

submitted to the EMBL Data Library, January 1995

A:Description: Sequence of a 37 kb DNA fragment from chromosome XII of Saccharomyces cer

A:Reference number: S50950

A:Accession: S50955

A:Molecule type: DNA

A:Residues: 1-116 <WED>

A:Cross-references: EMBL:Z47973; NID:G642313; PIDN:CAA87992.1; PID:G642319

R:Wedler, H.; Wedler, E.; Scharfe, M.; Wambutt, R.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64792

A:Accession: S64817

A:Molecule type: DNA

A:Residues: 1-116 <WED>

A:Cross-references: EMBL:Z73170; NID:G1360279; PIDN:CAA97519.1; PID:G1360280; MIPS:YLL06
A:Experimental source: strain S288C

```

C;Genetics:
A;Gene: SGD:GIN11
A;Cross-references: SGD:S0003988; MIPS:YLL065w
A;Map position: 12L

Query Match      69.2%; Score 36; DB 2; Length 116;
Best Local Similarity 75.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLHKVHFY 8
DB 46 FTHKVHHY 53

RESULT 3
PC1120
Glycoprotein-fucosylgalactoside alpha-N-acetylgalactosaminyltransferase (EC 2.4.1.40) A2
N;Alternate names: histo-blood group A2 transferase
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 03-Jun-2002
C;Accession: PC1120
R;Yamamoto, F.; McNeill, P.D.; Hakomori, S.
Biochem. Biophys. Res. Commun. 187, 366-374, 1992
A;Title: Human histo-blood group A2 transferase coded by A2 allele, one of the A subtype
the carboxyl terminal.
A;Reference number: PC1120; MUID:92392351; PMID:1520322
A;Accession: PC1120
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-295 <YAM>
A;Cross-references: GB:S44054; NID:g255066; PIDN:AA23167.1; PID:g255067
C;Superfamily: histo-blood group 1 transferase
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match      69.2%; Score 36; DB 2; Length 295;
Best Local Similarity 71.4%; Pred. No. 28;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
DB 65 HRVHYV 71

RESULT 4
A34933
Glycoprotein-fucosylgalactoside alpha-N-acetylgalactosaminyltransferase (EC 2.4.1.40) A1
N;Alternate names: alpha-3-N-acetylgalactosaminyltransferase; blood-group substance A-de
e; histo-blood group A glycosyltransferase
C;Species: Homo sapiens (man)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 19-Jul-2002
C;Accession: A34933; S09593; S13173; PC1164
R;Yamamoto, F.; Marken, J.; Tsuji, T.; White, T.; Clausen, H.; Hakomori, S.
J. Biol. Chem. 265, 1146-1151, 1990
A;Title: Cloning and characterization of DNA complementary to human UDP-GalNAc: Fucalpha
A;Reference number: A34933; MUID:90110098; PMID:2104828
A;Accession: A34933
A;Molecule type: mRNA
A;Residues: 1-353 <YAM1>
A;Cross-references: GB:J05175; NID:g340077; PIDN:AA36792.1; PID:g340078
R;Yamamoto, F.; Marken, J.; Clausen, H.; White, T.; Marken, J.; Hakomori, S.I.
Nature 345, 229-233, 1990
A;Title: Molecular genetic basis of the histo-blood group ABO system.
A;Reference number: S09593; MUID:90238543; PMID:2333095
A;Accession: S09593
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-353 <YAM2>
R;Navaratnam, N.; Findlay, J.B.C.; Keen, J.N.; Watkins, W.M.
Biochem. J. 271, 93-98, 1990
A;Title: Purification, properties and partial amino acid sequence of the blood-group-A-g
A;Reference number: S13173; MUID:91024951; PMID:2121135
A;Accession: S13173
A;Status: preliminary

```

```

A;Molecule type: protein
A;Residues: 'X',65-73,'IS',76-77 <NAV>
R;Kominato, Y.; McNeill, P.D.; Yamamoto, M.; Russell, M.; Hakomori, S.; Yamamoto, F.
Biochem. Biophys. Res. Commun. 189, 154-164, 1992
A;Title: Animal histo-blood group ABO genes.
A;Reference number: PC1164; MUID:93080551; PMID:1449469
A;Accession: PC1164
A;Molecule type: DNA
A;Residues: 145-154,'P',156-333 <KOM>
C;Comment: This enzyme forms group A blood type determinants from H antigen determinants
minor subtypes, for example A2 (see PIR:PC1120). Inactive alleles are responsible for g1
C;Genetics:
A;Gene: GDB:ABO
A;Cross-references: GDB:118956; OMIM:110300
A;Map position: 9q34.1-9q34.2
C;Complex: homodimer
C;Function:
A;Description: transfers N-acetylgalactosamine from UDP-N-acetylgalactosamine to glycopro
C;Superfamily: histo-blood group 1 transferase
C;Keywords: glycoprotein; glycosyltransferase; Golgi apparatus; hexosyltransferase; polyr
F.1-11/Domain: intracellular #status predicted <INT>
F.12-37/Domain: transmembrane #status predicted <TM>
F.38-353/Domain: trans-Golgi network luminal #status predicted <LUM>
F.112/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match      69.2%; Score 36; DB 1; Length 353;
Best Local Similarity 71.4%; Pred. No. 33;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
DB 144 HRVHYV 150

RESULT 5
T46126
hypothetical protein T2J13.120 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46126
R;Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Mayer, K
submitted to the Protein Sequence Database, November 1999
A;Reference number: Z23023
A;Accession: T46126
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-363 <RIE>
A;Cross-references: EMBL:AL132967
A;Experimental source: cultivar Columbia; BAC clone T2J13
C;Genetics:
A;Map position: 3
A;Introns: 5/2; 290/3
A;Note: T2J13.120

Query Match      69.2%; Score 36; DB 2; Length 363;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHFY 8
DB 168 LYKVHFY 174

RESULT 6
T46104
hypothetical protein T25B15.110 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 28-Jul-2000
C;Accession: T46104
R;Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.I
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23021
A;Accession: T46104

```

A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-412 <ALC>
 A;Cross-references: EMBL:AL132972
 A;Experimental source: cultivar Columbia; BAC clone T25B15
 C;Genetics:
 A;Map position: 3
 A;Introns: 20/3; 155/3; 220/3; 317/1; 369/1; 384/3
 A;Note: T25B15.110
 C;Superfamily: Arabidopsis thaliana hypothetical protein F24B22.230

Query Match 69.2%; Score 36; DB 2; Length 412;
 Best Local Similarity 75.0%; Pred. No. 39;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKVHFV 9
 |||||
 Db 128 LHKVSPYI 135

RESULT 7

S60622
 cytochrome-c oxidase [EC 1.9.3.1] chain I - brine shrimp mitochondrion
 C;Species: mitochondrion Artemia franciscana (brine shrimp)
 C;Date: 23-Feb-1996 #sequence_revision 26-Jul-1996 #text_change 07-Dec-1999
 C;Accession: S60622; S60638
 R;Valverde, J.R.
 submitted to the EMBL Data Library, June 1994
 A;Reference number: S60622
 A;Accession: S60622
 A;Molecule type: DNA
 A;Residues: 1-512 <VAL>
 A;Cross-references: EMBL:X69067; NID:G505262; PIDN:CAA48806.1; PID:G578437
 R;Perez, M.L.; Valverde, J.R.; Batuecas, B.; Amat, F.; Marco, R.; Garesse, R.
 J. Mol. Evol. 38, 156-168, 1994
 A;Title: Speciation in the Artemia genus: mitochondrial DNA analysis of bisexual and par
 A;Reference number: S60624; MUID:94233692; PMID:8169960
 A;Accession: S60638
 A;Molecule type: DNA
 A;Residues: 146-311 <PER>
 A;Cross-references: EMBL:X69067
 C;Genetics:
 A;Gene: COI
 A;Genome: COI
 A;Genetic code: SGC4
 C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as
 transmembrane protein
 F;9-455/Domain: cytochrome-c oxidase chain I homology <COI>
 F;59,376/Binding site: heme a iron (His) (axial ligands) #status predicted
 F;238,288,289/Binding site: copper (His) #status predicted
 F;238-242/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F;242/Binding site: oxygen (Tyr) #status predicted
 F;366/Binding site: magnesium (His) (shared with chain II) #status predicted
 F;374/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 69.2%; Score 36; DB 2; Length 512;
 Best Local Similarity 66.7%; Pred. No. 49;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLKVVHFV 9
 |||||
 Db 406 FLKVVHFI 414

RESULT 8

AH0655
 Probable regulatory protein STV1347 [imported] - Salmonella enterica subsp. enterica ser
 C;Species: Salmonella enterica subsp. enterica serovar Typhi
 A;Note: this species has also been called Salmonella typhi
 C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 19-Nov-2002
 C;Accession: AH0655
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
 A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AH0655
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-250 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD08427.1; PID:G16502470; GSPDB:GN00176
 C;Genetics:
 A;Gene: STV1347
 C;Superfamily: regulatory protein gutr

Query Match 67.3%; Score 35; DB 2; Length 250;
 Best Local Similarity 71.4%; Pred. No. 36;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLKVVHF 7
 |||||
 Db 153 FTHQVHF 159

RESULT 9

T32200
 hypothetical protein T02B11.4 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C;Accession: T32200
 R;Goela, D.

submitted to the EMBL Data Library, September 1997
 A;Description: The sequence of C. elegans cosmid T02B11.
 A;Reference number: Z21135

A;Accession: T32200
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-251 <GOE>
 A;Cross-references: EMBL:AF022979; PIDN:AAB69902.1; GSPDB:GN00023; CESP:T02B11.4
 A;Experimental source: strain Bristol N2; clone T02B11
 C;Genetics:
 A;Gene: CESP:T02B11.4
 A;Map position: 5
 A;Introns: 47/3; 69/1; 92/1; 180/2

Query Match 67.3%; Score 35; DB 2; Length 251;
 Best Local Similarity 62.5%; Pred. No. 36;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLKVVHFY 8
 |||||
 Db 226 FLYPIHFY 233

RESULT 10

S56162
 MDCR15 protein - human
 C;Species: Homo sapiens (man)
 C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
 C;Accession: S56162
 R;Barella, L.; Loetscher, M.; Tobler, A.; Baggiolini, M.; Moser, B.
 Biochem. J. 309, 773-779, 1995
 A;Title: Sequence variation of a novel heptahelical leucocyte receptor through alternati
 A;Reference number: S56162; MUID:95366951; PMID:7633692
 A;Accession: S56162
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-327 <BAR>
 A;Cross-references: EMBL:X68829; NID:G840783; PIDN:CAA48723.1; PID:G840784
 C;Superfamily: vertebrate rhodopsin

Query Match 67.3%; Score 35; DB 2; Length 327;
 Best Local Similarity 85.7%; Pred. No. 47;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHFY 8
|||||

Db 83 LHKVNPFY 89

RESULT 11
C71952
queine tRNA-ribosyltransferase - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 29-Sep-1999
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; PMID:99120557; PMID:9923682
A:Accession: C71952
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-371 <ARN>
A:Cross-references: GB:AE001464; GB:AE001439; NID:G4154789; PIDN:AAD05847.1; PID:G415479
A:Experimental source: strain J99
C:Genetics:
A:Gene: tsg
C:Superfamily: queine tRNA-ribosyltransferase

Query Match 67.3%; Score 35; DB 2; Length 371;
Best Local Similarity 62.5%; Pred. No. 54;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHFYV 9
|||

Db 333 LHNLFY 340

RESULT 12
A64555
tRNA-queine transglycosylase - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 29-Sep-1999
C:Accession: A64555
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; PMID:97394467; PMID:9252185
A:Accession: A64555
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-371 <TOM>
A:Cross-references: GB:AE000547; GB:AE000511; NID:G2313377; PIDN:AAD07350.1; PID:G231337
C:Superfamily: queine tRNA-ribosyltransferase

Query Match 67.3%; Score 35; DB 2; Length 371;
Best Local Similarity 62.5%; Pred. No. 54;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHFYV 9
|||

Db 333 LHNLFY 340

RESULT 13
S26667
G protein-coupled receptor BLR1 - human
C:Species: Homo sapiens (man)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C:Accession: S26667

R;Dobner, T.; Wolf, I.; Emrich, T.; Lipp, M.
Eur. J. Immunol. 22, 2795-2799, 1992
A:Title: Differentiation-specific expression of a novel G protein-coupled receptor from
A:Reference number: S26667; MUID:93049615; PMID:1425907
A:Accession: S26667
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-372 <DOB>
A:Cross-references: EMBL:X68149; NID:G29459; PIDN:CAA48252.1; PID:G29460
C:Genetics:
A:Gene: GDB:BLR1
A:Cross-references: GDB:136235; OMIM:601613
A:Map position: 15q26.1-15q26.1
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 67.3%; Score 35; DB 2; Length 372;
Best Local Similarity 85.7%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHFY 8
|||||

Db 128 LHKVNPFY 134

RESULT 14
F64419
hypothetical protein MJ0958 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 04-Mar-2000
C:Accession: F64419
R;Balt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: F64419
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-426 <BUL>
A:Cross-references: GB:U67539; GB:L77117; NID:G1591619; PIDN:AAB98971.1; PID:G1499795; T
C:Genetics:
C:Superfamily: Methanococcus jannaschii hypothetical protein MJ0958

Query Match 67.3%; Score 35; DB 2; Length 426;
Best Local Similarity 57.1%; Pred. No. 62;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
|||

Db 358 HKLHYI 364

RESULT 15
T03390
4-coumarate-CoA ligase (EC 6.2.1.12) isoform 2 - rice
N:Alternate names: 4-coumaroyl-CoA synthetase
C:Species: Oryza sativa (rice)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 03-Jun-2002
C:Accession: T03390
R;Zhao, Y.; Kung, S.D.; Bottino, P.J.
submitted to the EMBL Data Library, July 1995
A:Description: 4-coumarate: CoA ligase genes in rice: divergent structure and differenti
A:Reference number: Z14921
A:Accession: T03390
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-569 <ZHA>
A:Cross-references: EMBL:L43362; NID:G893293; PIDN:AAA69580.1; PID:G893294
C:Genetics:

2;Note: 4cl.2
C;Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology
C;Keywords: acid-thiol ligase; coenzyme A; flavonoid biosynthesis
F;89-562/Domain: acetate-CoA ligase homology <ACL>

Query Match 67.3%; Score 35; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKVHF 7
|
|
|
|
|
|
Db 540 LHKVHF 545

Search completed: December 16, 2003, 14:19:14
Job time : 11.3333 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:35 ; Search time 5.66667 Seconds
(without alignments)
74.689 Million cell updates/sec

Title: US-09-870-089b-5

Perfect score: 52

Sequence: 1 FLHKVHFYV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	40	76.9	515	1	MATK_HELBU
2	37	71.2	181	1	Y051 BPT4
3	37	71.2	515	1	MATK TRILU
4	37	71.2	515	1	MATK TRIUN
5	37	71.2	517	1	MATK TRIMC
6	37	71.2	689	1	L100 ADECC
7	36	69.2	354	1	EGAT HUMAN
8	36	69.2	505	1	MATK KUNCA
9	36	69.2	512	1	MATK KUNER
10	36	69.2	512	1	COX1 ARTSF
11	35	67.3	371	1	TGT_HELPJ
12	35	67.3	371	1	TGT_HELPY
13	35	67.3	372	1	CCR5 HUMAN
14	35	67.3	426	1	Y958 METJA
15	35	67.3	504	1	MATK ARUDI
16	35	67.3	569	1	4CL2 ORISA
17	34	65.4	319	1	PTE1 HUMAN
18	34	65.4	370	1	TRMU BUCBP
19	34	65.4	374	1	CCR5 MOUSE
20	34	65.4	374	1	CCR5 RAT
21	34	65.4	511	1	MATK ACCAL
22	33	63.5	153	1	MRGA BACSU
23	33	63.5	189	1	IMPI YEAST
24	33	63.5	237	1	GIDB STRMU
25	33	63.5	300	1	ER25 SCHPO
26	33	63.5	331	1	HFLC TREPA
27	33	63.5	432	1	DCDA BACMT
28	33	63.5	500	1	MATK BRASC
29	33	63.5	505	1	MATK NUPVA
30	33	63.5	507	1	MATK MICAL
31	33	63.5	509	1	MATK ARPGI
32	33	63.5	509	1	MATK NYMAD
33	33	63.5	512	1	MATK_ZANAE

ALIGNMENTS

RESULT 1
MATK_HELBU STANDARD; PRT; 515 AA.
AC Q9XPNG; 33 63.5 514 1 MATK_PHOOC
35 63.5 520 1 MATK_ASPEL
36 63.5 520 1 MATK_CONMU
37 63.5 626 1 RN17 MOUSE
38 63.5 716 1 PBPE BACSU
39 63.5 813 1 PHSG CHLMU
40 63.5 814 1 PHSG CHLTR
41 63.5 965 1 PSDI CAEEL
42 61.5 119 1 RS6 EUCBP
43 61.5 229 1 ARAD BACSU
44 61.5 318 1 LPSA BACNO
45 61.5 371 1 TGT_NEIMA

OS Helonias bullata (Swamp pink).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
OC Helonias.
OK NCBI_TaxID=50364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Kazempour Osaloo S., Utech F.H., Ohara M., Kawano S.;
RT "Molecular systematics of Trilliaceae I. Phylogenetic analyses of Trillium using matK gene sequences.";
RL J. Plant Res. 112:35-49(1999).
CC -!- FUNCTION: Probably assists in splicing chloroplast group II introns (By similarity).
CC -!- SIMILARITY: BELONGS TO THE INTRON MATURASE FAMILY 2. MATK SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB017375; BAA36791.1; --
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
DR mRNA processing; Chloroplast.
KW mRNA processing; Chloroplast.
SQ SEQUENCE 515 AA; 61545 MW; D6C14A5AB0DA468F CRC64;
Query Match 76.9%; Score 40; DB 1; Length 515;
Best Local Similarity 66.7%; Pred. No. 4.1;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 FLHKVHFYV 9
DB 247 FLERHFYV 255
RESULT 2
Y051 BPT4 STANDARD; PRT; 181 AA.
ID Y051 BPT4
AC P39240; Q96215;

Q9ghb2 phoenix dac
Q9tna5 aspidistra
Q9tnb1 canvallaria
Q99mv7 mus musculus
Q07868 bacillus su
Q9pke6 chlamydia m
O84250 chlamydia t
Q18115 caenorhabdi
Q89341 buchnera ap
P94525 bacillus su
P39907 bacteroides
Q9jva4 neisseria m

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AB017396; BAA36812.1; -.
 InterPro; IPR000442; Intron_maturase2.
 InterPro; IPR002866; MatK_N.
 Pfam; PF01348; Intron_maturas2; 1.
 Pfam; PF01824; MatK_N; 1.
 mRNA processing; Chloroplast.
 SEQUENCE 515 AA; 61074 MW; 6D750AF7F444DB50 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 515;
 Best Local Similarity 66.7%; Pred. NO. 15;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLHKVHFYV 9
 DB 247 FLERTHYFV 255

RESULT 4
 MATK_TRIUN STANDARD; PRT; 515 AA.
 ID Q9XPB8;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Maturase K (intron maturase).
 MATK.
 OS Trillium undulatum (Painted trillium).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Trilliaceae;
 OC Trillium.
 OX NCBI_TaxID=82504;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RA Kazempour Oesaloo S., Utech F.H., Ohara M., Kawano S.;
 RT Trillium using matk gene sequences.;
 RL J. Plant Res. 112:35-49(1999).
 CC -!- FUNCTION: Probably assists in splicing chloroplast group II
 CC introns (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE INTRON MATURASE FAMILY 2. MATK
 CC SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AB017413; BAA36829.1; -.
 InterPro; IPR000442; Intron_maturase2.
 InterPro; IPR002866; MatK_N.
 Pfam; PF01348; Intron_maturas2; 1.
 Pfam; PF01824; MatK_N; 1.
 mRNA processing; Chloroplast.
 SEQUENCE 515 AA; 61306 MW; 1CB6F63DFB45AED4 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 515;
 Best Local Similarity 66.7%; Pred. NO. 15;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 FLHKVHYV 9
Db 247 FLERTHYV 255

RESULT 5
MATK_TRIMC STANDARD; PRT; 517 AA.
AC Q9XPE1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maturase K (Intron maturase).
GN MATK
OS Trillium maculatum (Spotted waterbush).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Trilliales;
OC Trillium.
OX NCBI_TaxID=82490;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Kazempour Osaloo S., Utech F.H., Ohara M., Kawano S.;
RT "Molecular systematics of Trilliaceae I. Phylogenetic analyses of
RT Trillium using matK gene sequences.";
RL J. Plant Res. 112:35-49(1999).
CC -1- FUNCTION: Probably assists in splicing chloroplast group II
CC introns (By similarity).
CC -1- SIMILARITY: BELONGS TO THE INTRON MATURASE FAMILY 2. MATK
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB017397; BAA36813.1; -
DR InterPro; IPR000442; Intron_maturase2.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Trillium.
DR mRNA processing; Chloroplast.
SQ SEQUENCE 517 AA; 61378 MW; C18AB9643A7B5FED CRC64;

Query Match 71.2%; Score 37; DB 1; Length 517;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLHKVHYV 9
Db 247 FLERTHYV 255

RESULT 6
L100 ADECC STANDARD; PRT; 689 AA.
AC Q65957;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Late 100 kDa protein.
OS Canine adenovirus type 1 (strain CLL), and
OS Canine adenovirus type 1 (strain R1261).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=69150, 69151;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLL;
RA Campbell J.B., Zhao Y.;

Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=R1261;
RX MEDLINE=97275900; PubMed=9129661;
RA Morrison M.D., Onions D.E., Nicolson L.;
RT "Complete DNA sequence of canine adenovirus type 1.";
RL J. Gen. Virol. 78:873-878(1997).
CC -1- FUNCTION: THE 100 KDA PROTEIN IS A LATE NONSTRUCTURAL PROTEIN
CC INVOLVED IN TRANSPORT OF HEXON FROM CYTOPLASM TO THE NUCLEUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; US5001; AAB05446.1; -
DR EMBL; Y07760; CAA69039.1; -
DR InterPro; IPR003381; Adeno100.
DR Pfam; PF02438; adeno100; 1.
DR Late protein; Transport.
SQ SEQUENCE 689 AA; 77373 MW; 6062D58E0ACE7763 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 689;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HKVHFY 8
Db 578 HKIHFY 583

RESULT 7
BGAT_HUMAN STANDARD; PRT; 354 AA.
ID BGAT_HUMAN
AC P16442; Q14490;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Histo-blood group ABO system transferase (NAGAT) [Includes:
DE Glycoprotein-fucosylgalactoside alpha-N-
DE acetylglucosaminyltransferase (EC 2.4.1.40) (Fucosylglycoprotein
DE alpha-N-acetylglucosaminyltransferase) (Histo-blood group A
DE transferase) (A transferase); Glycoprotein-fucosylgalactoside alpha-
DE galactosyltransferase (EC 2.4.1.37) (Fucosylglycoprotein 3-alpha-
DE galactosyltransferase) (Histo-blood group B transferase) (B
DE transferase)].
GN ABO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90110098; PubMed=2104828;
RA Yamamoto F.-I., Marken J., Tsuji T., White T., Clausen H.,
RA Hakomori S.-I.;
RT "Cloning and characterization of DNA complementary to human UDP-
RT GalNAc: Fuc alpha 1-->2Gal alpha 1-->3GalNAc transferase (histo-blood
RT group A transferase) mRNA.";
RL J. Biol. Chem. 265:1146-1151(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90238543; PubMed=2333095;
RA Yamamoto F.-I., Clausen H., White T., Marken J., Hakomori S.-I.;
RT "Molecular genetic basis of the histo-blood group ABO system.";
RL Nature 345:229-233(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95118355; PubMed=7598760;

```


RA Bennett E.P., Steffensen R., Clausen H., Weghuis D.O.,
 RA Geurts van Kessel A.;
 RA "Genomic cloning of the human histo-blood group ABO locus.";
 RL Biochem. Biophys. Res. Commun. 206:318-325(1995).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=95298046; PubMed=7779106;
 RX Bennett E.P., Steffensen R., Clausen H., Weghuis D.O.,
 RA Geurts van Kessel A.;
 RA "Genomic cloning of the human histo-blood group ABO locus.";
 RL Biochem. Biophys. Res. Commun. 211:347-347(1995).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP Yamamoto F.-I.;
 RA "Human histo-blood group ABO gene locus alleles.";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN CHARACTERIZATION.
 RP MEDLINE=91035461; PubMed=2121736;
 RX Yamamoto F.-I., Hakomori S.-I.;
 RA "Sugar-nucleotide donor specificity of histo-blood group A and B
 RT transferases is based on amino acid substitutions.";
 RL J. Biol. Chem. 265:19257-19262(1990).
 CC -!- FUNCTION: This protein is the basis of the ABO blood group system.
 CC The histo-blood group ABO involves three carbohydrate antigens: A,
 CC B, and H. A, B, and AB individuals express a glycosyltransferase
 CC activity that converts the H antigen to the A antigen (by addition
 CC of UDP-GalNAc) or to the B antigen (by addition of UDP-Gal),
 CC whereas O individuals lack such activity.
 CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + glycoprotein-
 CC alpha-L-fucosyl-(1-2)-D-galactose = UDP + glycoprotein-N-acetyl-
 CC alpha-D-galactosaminyl-(1,3)-[alpha-L-fucosyl-(1,2)]-D-galactose.
 CC -!- CATALYTIC ACTIVITY: UDP-galactose + alpha-L-fucosyl-(1->2)-D-
 CC galactosyl-R = UDP + alpha-D-galactosyl-(1->3)-[alpha-L-
 CC fucosyl(1->2)]-D-galactosyl-R.
 CC -!- PATHWAY: Glycosylation.
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
 CC FORM IN TRANS CISTERNAE OF GOLGI. SOLUBLE FORM IN BODY FLUIDS.
 CC -!- PTM: THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY
 CC PROTEOLYTIC PROCESSING.
 CC -!- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF THE A TRANSFERASE. THE
 CC B FORM DIFFERS BY A FEW RESIDUES SUBSTITUTIONS, THE O PHENOTYPE IS
 CC RESULT OF A SINGLE BASE FRAMESHIFT DELETION IN THE N-TERMINAL
 CC EXTREMITY OF THE GENE.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 6.

DR GO:0006486; P:protein amino acid glycosylation; NAS.
 DR InterPro; IPR005076; Glyco transf 6.
 DR Pfam; PF03414; Glyco transf 6; 1.
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
 KW Signal-anchor; Golgi stack; Polymorphism; Blood group antigen;
 KW 3D-structure.
 FT CHAIN 54 354 FUCOSYLGLYCOPROTEIN ALPHA-N-
 FT ACETYLGLACTOSAMINYLTRANSFERASE, SOLUBLE
 FT FORM.
 FT DOMAIN 1 32
 FT TRANSMEM 33 53 CYTOLASMIC (POTENTIAL).
 FT SIGNAL-ANCHOR {TYPE-II MEMBRANE PROTEIN}
 FT (POTENTIAL).
 FT DOMAIN 54 354 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 80 81 CR -> W.
 FT VARIANT 156 156 /FTid=VAR_003408.
 FT VARIANT 176 176 P -> L (IN ALLELE A2).
 FT VARIANT 235 235 /FTid=VAR_003409.
 FT VARIANT 266 266 R -> G (IN GROUP B TRANSFERASE).
 FT VARIANT 268 268 G -> S (IN GROUP B TRANSFERASE).
 FT VARIANT 268 268 L -> M (IN GROUP B TRANSFERASE).
 FT VARIANT 268 268 /FTid=VAR_003411.
 FT VARIANT 268 268 /FTid=VAR_003412.
 FT VARIANT 352 352 G -> A (IN GROUP B TRANSFERASE).
 FT VARIANT 352 352 /FTid=VAR_003413.
 FT VARIANT 352 352 R -> W (IN ALLELE B3 OF GROUP B
 FT TRANSFERASE).
 FT /FTid=VAR_003414.
 SQ SEQUENCE 354 AA; 40934 MW; A03DA16E630C1608 CRC64;
 Query Match 69.2%; Score 36; DB 1; Length 354;
 Best Local Similarity 71.4%; Pred. No. 15;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 3 HKVHFV 9
 Db 145 HRVHYV 151
 RESULT 8
 MATK_KUNCA
 ID MATK_KUNCA STANDARD; PRT; 505 AA.
 AC O9TKB9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Maturase K (Intron maturase).
 GN MATK.
 OS Kunzea capitata (Pink kunzea).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Myrtales; Myrtaceae; Kunzea.
 OX NCBI_TaxID=106043;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'Brien M.M., Quinn C.J., Wilson P.G.;
 RT "Molecular Systematics of the Leptospermum suballiance (Myrtaceae).";
 RL Aust. J. Bot. 48:621-628(2000).
 CC -!- FUNCTION: Probably assists in splicing chloroplast group II
 CC introns (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE INTRON MATURASE FAMILY 2. MATK
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF194723; AAF05930.1; -.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW mRNA processing; Chloroplast.
SQ SEQUENCE 505 AA; 60711 MW; 41B5EC246DC096A4 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 505;
Best Local Similarity 55.6%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 0; Gaps 0;

QY 1 FLHKVHFYV 9
Db 238 FFERIHFFV 246

RESULT 9
MATK_KUNER
ID MATK_KUNER STANDARD; PRT; 505 AA.
AC Q9TK58;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Maturase K (Intron maturase).
GN MATK.
OS Kunzea ericoides (Kanuka).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Myrtales; Myrtaceae; Kunzea.
OX NCBI_TaxID=106044;
RN [1]
RP SEQUENCE FROM N.A.
RA O'Brien M.M., Quinn C.J., Wilson P.G.;
RT "Molecular systematics of the Leptospermum suballiance (Myrtaceae).";
RL Aust. J. Bot. 48:621-628(2000).
CC -!- FUNCTION: Probably assists in splicing chloroplast group II
CC introns (By similarity).
CC -!- SIMILARITY: BELONGS TO THE INTRON MATURASE FAMILY 2. MATK
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF184724; AAF05931.2; -.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW mRNA processing; Chloroplast.
SQ SEQUENCE 505 AA; 60657 MW; 20D2B47963263B8D CRC64;

Query Match 69.2%; Score 36; DB 1; Length 505;
Best Local Similarity 55.6%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 0; Gaps 0;

QY 1 FLHKVHFYV 9
Db 238 FFERIHFFV 246

RESULT 10
COXI_ARTSF
ID COXI_ARTSF STANDARD; PRT; 512 AA.
AC Q37705;

```

```

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN COI OR CO-I.
OS Artemia sanfranciscana (Brine shrimp) (Artemia franciscana).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;
OC Artemiidae; Artemia.
OX NCBI_TaxID=6661;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94223692; PubMed=8169960;
RA Perez M.I., Valverde J.R., Batuecas B., Amat F., Marco R., Garesse R.;
RT "Speciation in the Artemia genus: mitochondrial DNA analysis of
RT bisexual and parthenogenetic brine shrimps.";
RL J. Mol. Evol. 38:156-168(1994).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 ferrocycytochrome c + O(2) = 4 ferrocycytochrome
CC c + 2 H(2)O.
CC -!- PATHWAY: Respiratory chain; terminal step.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69067; CAA48806.1; -.
DR PIR; S60622; S60622.
DR HSSP; P00396; ZOCC.
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASEI.
DR PROSITE; PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
FT METAL 59 59 IRON (HEME A AXIAL LIGAND) (PROBABLE).
FT METAL 238 238 COPPER B (PROBABLE).
FT METAL 242 242 COPPER B (PROBABLE).
FT METAL 288 288 COPPER B (PROBABLE).
FT METAL 289 289 COPPER B (PROBABLE).
FT METAL 374 374 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
FT METAL 376 376 IRON (HEME A AXIAL LIGAND) (PROBABLE).
FT CROSSLINK 238 242 1'-histidyl-3'-tyrosine (By similarity).
SQ SEQUENCE 512 AA; 56491 MW; 4AA1E1BB64913E3 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 512;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFYV 9
Db 406 FLKVVHFFI 414

RESULT 11
TGT_HELPJ
ID TGT_HELPJ STANDARD; PRT; 371 AA.
AC Q9ZMF4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

28-FEB-2003 (Rel. 41, Last annotation update)
 Queuine tRNA-ribosyltransferase (EC 2.4.2.29) (tRNA-guanine
 transglycosylase) (Guanine insertion enzyme).
 TGT OR JHP0266.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 ON NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deLonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 CC -!- FUNCTION: Exchanges the guanine residue with 7-aminomethyl-7-
 CC deazaguanine in tRNAs with GU(N) anticodons (tRNA-Asp, -Asn, -His
 CC and -Tyr). After this exchange, a cyclopentendiol moiety is
 CC attached to the 7-aminomethyl group of 7-deazaguanine, resulting
 CC in the hypermodified nucleoside queuine (Q) (7-(((4,5-cis-
 CC dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanosine) (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: tRNA guanine + queuine = tRNA queuine +
 CC guanine.
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE QUEUINE TRNA-RIBOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE001464; AAC05847.1; -;
 CC PIR; C71952; C71952.
 CC HSP; P28720; 1PUD.
 CC HAMAP; MF 00168; -; 1.
 CC InterPro; IPR004803; Qcrna_ribo_trans.
 CC InterPro; IPR002616; tRNA_ribo_trans.
 CC Pfam; PF01702; TGT; 1.
 CC TIGRFAMs; TIGR00430; Q crna_tgt; 1.
 CC TIGRFAMs; TIGR00449; tgt_general; 1.
 CC Queuosine biosynthesis; Transferrase; Glycosyltransferase;
 CC tRNA processing; Zinc; Complete proteome.
 CC ACT_SITE 91 91 BY SIMILARITY.
 CC METAL 266 266 BY SIMILARITY.
 CC METAL 303 303 ZINC (BY SIMILARITY).
 CC METAL 305 305 ZINC (BY SIMILARITY).
 CC METAL 308 308 ZINC (BY SIMILARITY).
 CC METAL 334 334 ZINC (BY SIMILARITY).
 CC METAL 371 AA; 41506 MW; 155668901BB7FC3D CRC64;
 SQ SEQUENCE 371 AA; 41506 MW; 155668901BB7FC3D CRC64;
 Query Match 67.3%; Score 35; DB 1; Length 371;
 Best Local Similarity 62.5%; Pred. No. 25;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LHKVHPV 9
 |||:||||
 Db 333 LHNLFHYL 340
 RESULT 12
 TGT_HELPY
 ID -TGT_HELPY STANDARD; PRT; 371 AA.
 AC O08314;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)
 Queuine tRNA-ribosyltransferase (EC 2.4.2.29) (tRNA-guanine
 transglycosylase) (Guanine insertion enzyme).
 TGT OR HP0281.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 ON NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-P., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams W.D., Hickley E.K., J.M.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori.";
 RL Nature 388:539-547(1997).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC STRAIN=PI;
 CC Bereswill S., Fassbinder F., Voelzing C., Haas R., Kist M.;
 CC Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Exchanges the guanine residue with 7-aminomethyl-7-
 CC deazaguanine in tRNAs with GU(N) anticodons (tRNA-Asp, -Asn, -His
 CC and -Tyr). After this exchange, a cyclopentendiol moiety is
 CC attached to the 7-aminomethyl group of 7-deazaguanine, resulting
 CC in the hypermodified nucleoside queuine (Q) (7-(((4,5-cis-
 CC dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanosine) (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: tRNA guanine + queuine = tRNA queuine +
 CC guanine.
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE QUEUINE TRNA-RIBOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000547; AAC07350.1; -;
 CC PIR; Y12061; CAA72784.1; -;
 CC HSP; P28720; 1PUD.
 CC TIGR; HP0281; -; 1.
 CC HAMAP; MF 00168; -; 1.
 CC InterPro; IPR004803; Qcrna_ribo_trans.
 CC InterPro; IPR002616; tRNA_ribo_trans.
 CC Pfam; PF01702; TGT; 1.
 CC TIGRFAMs; TIGR00430; Q crna_tgt; 1.
 CC TIGRFAMs; TIGR00449; tgt_general; 1.
 CC Queuosine biosynthesis; Transferrase; Glycosyltransferase;
 CC tRNA processing; Zinc; Complete proteome.
 CC ACT_SITE 91 91 BY SIMILARITY.
 CC METAL 266 266 BY SIMILARITY.
 CC METAL 303 303 ZINC (BY SIMILARITY).
 CC METAL 305 305 ZINC (BY SIMILARITY).
 CC METAL 308 308 ZINC (BY SIMILARITY).
 CC METAL 334 334 ZINC (BY SIMILARITY).
 CC METAL 10 11 NN -> KH (IN STRAIN PI).
 CC VARIANT 18 18 N -> D (IN STRAIN PI).
 CC VARIANT 46 46 A -> V (IN STRAIN PI).
 CC VARIANT 48 48 E -> G (IN STRAIN PI).
 CC VARIANT 72 73 EE -> GQ (IN STRAIN PI).

```

FT VARIANT 76 76 G -> V (IN STRAIN P1).
FT VARIANT 79 79 R -> H (IN STRAIN P1).
FT VARIANT 84 84 Y -> Q (IN STRAIN P1).
FT VARIANT 108 108 D -> G (IN STRAIN P1).
FT VARIANT 119 120 SK -> NN (IN STRAIN P1).
FT VARIANT 169 169 M -> L (IN STRAIN P1).
FT VARIANT 177 177 K -> N (IN STRAIN P1).
FT VARIANT 181 181 S -> N (IN STRAIN P1).
FT VARIANT 206 206 E -> K (IN STRAIN P1).
FT VARIANT 233 233 A -> T (IN STRAIN P1).
FT VARIANT 258 258 S -> G (IN STRAIN P1).
FT VARIANT 304 304 A -> T (IN STRAIN P1).
SQ SEQUENCE 371 AA; 41415 MW; 2F579B19EA90FD6A CRC64;

Query Match 67.3%; Score 35; DB 1; Length 371;
Best Local Similarity 82.5%; Pred. No. 25;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKVHFVY 9
DB 333 LHNLFVYL 340

RESULT 13
CCRS5 HUMAN STANDARD; PRT; 372 AA.
ID C32302; Q14811;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE C-X-C chemokine receptor type 5 (CXCR-5) (CXCR-5) (Burkitt's lymphoma
DE receptor 1) (Monocyte-derived receptor 15) (MDR15).
GN BLR1 OR CXCR5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;

[1]
RN SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Lymphocytes;
RC MEDLINE=93049615; PubMed=1425907;
RA Dobner T., Wolf I., Enrich T., Lipp M.;
RT "differentiation-specific expression of a novel G protein-coupled
RT receptor from Burkitt's lymphoma."
RL Eur. J. Immunol. 22:2795-2799(1992).
[2]
RN SEQUENCE FROM N.A. (ISOFORM SHORT).
RC TISSUE=Blood;
RC MEDLINE=95366951; PubMed=7639692;
RA Barella L., Loetscher M., Tobler A., Baggiolini M., Moser B.;
RT "Sequence variation of a novel heptahelical leucocyte receptor
RT through alternative transcript formation."
RL Biochem. J. 309:773-779(1995).
[3]
RN LIGAND BINDING.
RC MEDLINE=98130629; PubMed=9463416;
RA Legler D.F., Loetscher M., Stuber Roos R., Clark-Lewis I.,
RA Baggiolini M., Moser B.;
RT "B cell-attracting chemokine 1, a human CXC chemokine expressed in
RT lymphoid tissues, selectively attracts B lymphocytes via
RT BLR1/CXCR5."
RL J. Exp. Med. 187:655-660(1998).
[4]
RN FUNCTION: CYTOKINE RECEPTOR THAT BINDS TO BLC. BLR1 EXERTS
RN POSSIBLY A REGULATORY FUNCTION IN BURKITT LYMPHOMA (BL)
RN LYMPHOMAGENESIS AND/OR B-CELL DIFFERENTIATION. IT IS A POTENTIAL
RN CANDIDATE FOR CELL-CELL INTERACTION, AND ACTIVATION OF MATURE B-
RN LYMPHOCYTES IN LYMPHATIC TISSUES.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P32302-1; Sequence=Displayed;
CC Name=Short;

```

```

CC IsoId=P32302-2; Sequence=VSP_001892;
CC -!- TISSUE SPECIFICITY: EXPRESSION IN MATURE B-CELLS AND BURKITT
CC LYMPHOMA CELLS.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC
CC EMBL; X68149; CAA48252.1; --
CC EMBL; X68829; CAA48723.1; --
CC PIR; S26667; S26667.
CC DR Genew; HGNC:1060; BLR1.
CC DR MIM; 601613; --
CC DR GO; GO:0005887; C: integral to plasma membrane; TAS.
CC DR GO; GO:0006928; P: cell motility; TAS.
CC DR GO; GO:0007186; P: G-protein coupled receptor protein signalin...; TAS.
CC DR InterPro; IPR000276; GPCR_Rhodopsn.
CC DR Pfam; PF00001; 7tm_1; 1.
CC DR PRINTS; PR00237; GPCR_Rhodopsn.
CC DR PROSITE; PS00237; G PROTEIN RECP F1.1; 1.
CC DR PROSITE; PS0262; G PROTEIN RECP F1.2; 1.
CC KW G-protein coupled receptor; Transmembrane; Glycoprotein; B-cell;
CC Alternative splicing; Polymorphism.
FT DOMAIN 1 55 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 56 76 1 (POTENTIAL).
FT DOMAIN 77 88 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 89 109 2 (POTENTIAL).
FT DOMAIN 110 124 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 125 145 3 (POTENTIAL).
FT DOMAIN 146 167 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 168 188 4 (POTENTIAL).
FT DOMAIN 189 219 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 220 240 5 (POTENTIAL).
FT DOMAIN 241 259 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 260 280 6 (POTENTIAL).
FT DOMAIN 281 304 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 305 325 7 (POTENTIAL).
FT DOMAIN 326 372 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 28 28 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC... ) (POTENTIAL).
FT DISULFID 122 202 POTENTIAL.
FT VARSPLIC 1 45 Missing (in isoform Short).
FT VARIANT 344 344 /FTId=VSP_001892.
FT G -> S (IN dBSNP:665648).
FT /FTId=VAR_011838.
SQ SEQUENCE 372 AA; 41955 MW; 6DF84C939492ACCF CRC64;

Query Match 67.3%; Score 35; DB 1; Length 372;
Best Local Similarity 85.7%; Pred. No. 25;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKVHFVY 8
DB 128 LHKVNFY 134

RESULT 14
Y958 METJA
ID Y958 METJA STANDARD; PRT; 426 AA.
AC O58358;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0958 precursor.
GN MJ0958.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.

```

```

OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervajave A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jamaeschii.";
RL Science 273:1058-1073 (1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67539; AAB98971.1; -
DR PIR; F64419; F64419.
DR TIGR; MJ0958; -.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 23 POTENTIAL
FT CHAIN 24 426 HYPOTHETICAL PROTEIN MJ0958.
SQ SEQUENCE 426 AA; 50694 MW; 25317BE81E203A99 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 426;
Best Local Similarity 57.1%; Pred. No. 28;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKVHFYV 9
DB 358 HKLHYI 364

RESULT 15
MATK_ARUDI
ID MATK_ARUDI STANDARD; PRT; 504 AA.
AC Q8W0R4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maturase K (Intron maturase).
GN MATK.
OS Aruncus dioicus (Goat's beard).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Rosales; Rosaceae; Rosaceae incertae sedis; Aruncus.
OX NCBI_TaxID=32220;
RN [1]
RP SEQUENCE FROM N.A.
RA Potter D., Gao F., Oh S.-H., Baggett S.;
RT "Phylogenetic relationships among putative genes encoding
RT polysialacturonase inhibitor proteins (PGIPs) in Rosaceae.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probably assists in splicing chloroplast group II
CC introns (By similarity).
CC -1- SIMILARITY: BELONGS TO THE INTRON MATURASE FAMILY 2. MATK
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF288094; AAL35988.1; -
DR InterPro; IPR000442; Intron_maturase2.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 2.
KW mRNA processing; Chloroplast.
SQ SEQUENCE 504 AA; 60201 MW; 5C59D4117987C05E CRC64;

Query Match 67.3%; Score 35; DB 1; Length 504;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLHKVHFY 8
DB 238 FFEKHFY 245

Search completed: December 16, 2003, 14:15:14
Job time : 6.66667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:09:45 ; Search time 25 seconds

(without alignments)
92.899 Million cell updates/sec

Title: US-09-870-089b-5

Perfect score: 52

Sequence: 1 FLHKVHFYV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacterioph.*
17: sp_archaeop.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	82.7	1043	Q8NLT2	Q8nlt2 corynebacte
2	40	76.9	515	Q9GHD1	Q9ghd1 helionopsis
3	40	76.9	515	Q9GHD5	Q9ghd5 helionopsis
4	40	76.9	515	Q9GHC9	Q9ghc9 helionopsis
5	40	76.9	515	Q9GHC8	Q9ghc8 helionopsis
6	40	76.9	515	Q9GHD2	Q9ghd2 helionopsis
7	40	76.9	515	Q9GHC5	Q9ghc5 chionograph
8	40	76.9	515	Q9GHD3	Q9ghd3 helionopsis
9	40	76.9	515	Q9GHD6	Q9ghd6 helionopsis
10	40	76.9	515	Q9GHD0	Q9ghd0 helionopsis
11	40	76.9	515	Q9GHD4	Q9ghd4 helionopsis
12	40	76.9	515	Q9GHD8	Q9ghd8 ypsilandra
13	40	76.9	515	Q9GHD7	Q9ghd7 helionopsis
14	40	76.9	515	Q9GHC7	Q9ghc7 chamaeliriu
15	40	76.9	515	Q8HV35	Q8hv35 pyrgophyllu
16	40	76.9	517	Q9GHC6	Q9ghc6 chionograph

17 39 75.0 209 4 Q9BT16
18 39 75.0 209 11 Q9CWL2
19 39 75.0 716 5 Q9VHM2
20 39 75.0 942 5 Q96OB2
21 39 75.0 1044 4 Q8WX49
22 39 75.0 1044 4 Q8NXC6
23 38 73.1 212 8 Q9XLP0
24 38 73.1 212 8 Q9XLN5
25 38 73.1 402 2 Q8KWR1
26 38 73.1 511 8 Q8HV19
27 38 73.1 511 8 Q8HV18
28 38 73.1 1807 5 Q8I389
29 37 71.2 71 5 Q8IAY5
30 37 71.2 311 5 Q9GS12
31 37 71.2 354 2 Q8KQW4
32 37 71.2 450 8 Q8HRJ6
33 37 71.2 450 8 Q8HRJ5
34 37 71.2 450 8 Q8HRJ4
35 37 71.2 450 8 Q8HRJ3
36 37 71.2 450 8 Q8HRJ2
37 71.2 513 8 Q9XPP3
38 37 71.2 515 8 Q9XR23
39 37 71.2 515 8 Q9XPN8
40 37 71.2 515 8 Q9GHC4
41 41 71.2 515 8 Q9XR22
42 37 71.2 515 8 Q9XPP4
43 37 71.2 515 8 Q9XPP0
44 37 71.2 515 8 Q9XR21
45 37 71.2 515 8 Q9XR24

ALIGNMENTS

RESULT 1
Q8NLT2 PRELIMINARY; PRT; 1043 AA.
AC Q8NLT2, 2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical membrane protein Cgl2855.
GN CGL2855.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
RX NCBI_TaxID=1718;
RV [1]
PF SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005283; BAC00249.1; --
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1043 AA; 112335 MW; 2A44C0E5431A29FC CRC64;

Query Match 82.7%; Score 43; DB 16; Length 1043;
Best Local Similarity 77.8%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLHKVHFYV 9
||| ||||
Db 22 FLHSAHFYV 30

RESULT 2
Q9GHD1 PRELIMINARY; PRT; 515 AA.
ID Q9GHD1
AC Q9GHD1; (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

```

DE Intron maturase (Maturase K).
GN MATK.
OS Heloniopsis orientalis.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
OC Heloniopsis.
OX NCBI_TaxID=87628;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=yellow flower;
RA Fuse S., Tamura M.N.;
RT "A phylogenetic analysis of the plastid matk gene with emphasis on
RT Melanthiaceae sensu lato.";
RL Plant Biol. 2:415-427(2000).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: AB040192; BAB16800.1; -.
DR InterPro: IPR000442; Intron_maturase2.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
KW mRNA processing; Chloroplast.
SQ SEQUENCE 515 AA; 61747 MW; 66A8A11DDA39BFBB CRC64;

Query Match 76.9%; Score 40; DB 8; Length 515;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFYV 9
DB 247 FLERIHFYV 255

RESULT 3
Q9GHD5 PRELIMINARY; PRT; 515 AA.
AC Q9GHD5;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Intron maturase (Maturase K).
GN MATK.
OS Heloniopsis orientalis.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
OC Heloniopsis.
OX NCBI_TaxID=87628;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=pink flower;
RA Fuse S., Tamura M.N.;
RT "A phylogenetic analysis of the plastid matk gene with emphasis on
RT Melanthiaceae sensu lato.";
RL Plant Biol. 2:415-427(2000).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: AB040194; BAB16802.1; -.
DR InterPro: IPR000442; Intron_maturase2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
KW mRNA processing; Chloroplast.
SQ SEQUENCE 515 AA; 61866 MW; 495C06AD09F7FE66 CRC64;

Query Match 76.9%; Score 40; DB 8; Length 515;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFYV 9
DB 247 FLERIHFYV 255

RESULT 4
Q9GHC9 PRELIMINARY; PRT; 515 AA.
AC Q9GHC9;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Intron maturase (Maturase K).
GN MATK.
OS Heloniopsis orientalis.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
OC Heloniopsis.
OX NCBI_TaxID=87628;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=pink flower;
RA Fuse S., Tamura M.N.;
RT "A phylogenetic analysis of the plastid matk gene with emphasis on
RT Melanthiaceae sensu lato.";
RL Plant Biol. 2:415-427(2000).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: AB040194; BAB16802.1; -.
DR InterPro: IPR000442; Intron_maturase2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
KW mRNA processing; Chloroplast.
SQ SEQUENCE 515 AA; 61866 MW; 495C06AD09F7FE66 CRC64;

Query Match 76.9%; Score 40; DB 8; Length 515;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFYV 9
DB 247 FLERIHFYV 255

RESULT 5
Q9GHC8 PRELIMINARY; PRT; 515 AA.
AC Q9GHC8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Intron maturase (Maturase K).
GN MATK.
OS Heloniopsis umbellata.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
OC Heloniopsis.
OX NCBI_TaxID=120002;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=pink flower;
RA Fuse S., Tamura M.N.;
RT "A phylogenetic analysis of the plastid matk gene with emphasis on
RT Melanthiaceae sensu lato.";
RL Plant Biol. 2:415-427(2000).

```

CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY)
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 DR EMBL; AB040195; BAB16803.1; -.
 DR InterPro; IPR000442; Intron_maturase2.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61945 MW; BB50E0A82BDD1F88 CRC64;
 Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHPYV 9
 DB 247 FLERIHFYV 255
 RESULT 6
 Q9GHD2 PRELIMINARY; PRT; 515 AA.
 ID Q9GHD2
 AC Q9GHD2;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Intron maturase (Maturase K).
 GN MATK.
 OS Helionopsis orientalis.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Helionopsis.
 OX NCBI_TaxID=87628;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=pink flower;
 RA Fuse S., Tamura M.N.;
 RT "A phylogenetic analysis of the plastid matk gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 DR EMBL; AB040191; BAB16799.1; -.
 DR InterPro; IPR000442; Intron_maturase2.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61780 MW; A4932FD7A8600BA8 CRC64;
 Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHPYV 9
 DB 247 FLERIHFYV 255
 RESULT 7
 Q9GHC5 PRELIMINARY; PRT; 515 AA.
 ID Q9GHC5
 AC Q9GHC5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Intron maturase (Maturase K).
 GN MATK.
 OS Chionographis japonica.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Chionographis.
 OX NCBI_TaxID=119999;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fuse S., Tamura M.N.;
 RT "A phylogenetic analysis of the plastid matk gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 DR EMBL; AB040198; BAB16806.1; -.
 DR InterPro; IPR000442; Intron_maturase2.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61353 MW; 998164C7FB40C2B1 CRC64;
 Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHPYV 9
 DB 247 FLERIHFYV 255
 RESULT 8
 Q9GHD3 PRELIMINARY; PRT; 515 AA.
 ID Q9GHD3
 AC Q9GHD3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Intron maturase (Maturase K).
 GN MATK.
 OS Helionopsis orientalis.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Helionopsis.
 OX NCBI_TaxID=87628;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=white flower;
 RA Fuse S., Tamura M.N.;
 RT "A phylogenetic analysis of the plastid matk gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 DR EMBL; AB040190; BAB16798.1; -.
 DR InterPro; IPR000442; Intron_maturase2.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61919 MW; 556C6A17A69204FF CRC64;
 Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFVY 9
 || :|||
 247 FLERIHFVY 255

RESULT 9
 Q9GHD6 PRELIMINARY; PRT; 515 AA.
 ID Q9GHD6;
 AC Q9GHD6;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Intron maturase (Maturase K).
 GN MATK.

OS Heloniopsis leucantha.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Heloniopsis.
 OX NCBI_TaxID=120001;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Fuse S., Tamura M.N.;
 RT "A phylogenetic analysis of the plastid matK gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC - FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC - INTRONS (BY SIMILARITY).
 CC - SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 DR EMBL; AB040187; BAB16795.1; --
 DR InterPro; IPR000442; Intron_maturase2.
 DR InterPro; IPR002866; MatK_N_
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61913 MW; 60674175DF2819DC CRC64;

Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFVY 9
 || :|||
 247 FLERIHFVY 255

RESULT 10
 Q9GHD0 PRELIMINARY; PRT; 515 AA.
 ID Q9GHD0;
 AC Q9GHD0;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Intron maturase (Maturase K).
 GN MATK.

OS Heloniopsis orientalis.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Heloniopsis.
 OX NCBI_TaxID=87628;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Fuse S., Tamura M.N.;
 RT "A phylogenetic analysis of the plastid matK gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC - FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II

CC INTRONS (BY SIMILARITY).
 CC - SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 DR EMBL; AB040193; BAB16801.1; --
 DR InterPro; IPR000442; Intron_maturase2.
 DR InterPro; IPR002866; MatK_N_
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61844 MW; 1BBFF413F5DE7848 CRC64;

Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFVY 9
 || :|||
 247 FLERIHFVY 255

RESULT 11
 Q9GHD4 PRELIMINARY; PRT; 515 AA.
 ID Q9GHD4;
 AC Q9GHD4;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Intron maturase (Maturase K).
 GN MATK.

OS Heloniopsis orientalis.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Heloniopsis.
 OX NCBI_TaxID=87628;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Fuse S., Tamura M.N.;
 RT "A phylogenetic analysis of the plastid matK gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC - FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC - INTRONS (BY SIMILARITY).
 CC - SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 DR EMBL; AB040189; BAB16797.1; --
 DR InterPro; IPR000442; Intron_maturase2.
 DR InterPro; IPR002866; MatK_N_
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61916 MW; 7EB6D08388182890 CRC64;

Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFVY 9
 || :|||
 247 FLERIHFVY 255

RESULT 12
 Q9GHD8 PRELIMINARY; PRT; 515 AA.
 ID Q9GHD8;
 AC Q9GHD8;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Intron maturase (Maturase K).

GN MATK.
 OS Ypsilandra thibetica.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Ypsilandra.
 OX NCBI_TaxID=120020;
 [1]
 RN SEQUENCE FROM N.A.
 RP Fuse S., Tamura M.N.;
 RA "A phylogenetic analysis of the plastid matK gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 CC EMBL; AB040185; BAB16793.2; --
 DR InterPro; IPR000442; Intron_maturase2.
 DR InterPro; IPR002866; MatK_N.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61796 MW; 1C8FFF71CAF6CE8 CRC64;
 Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHFV 9
 DB 247 FLERIHFV 255
 RESULT 13
 Q9GHD7 PRELIMINARY; PRT; 515 AA.
 AC Q9GHD7;
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Intron maturase (Maturase K).
 GN MATK.
 OS Heloniopsis kawaiol.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Heloniopsis.
 OX NCBI_TaxID=120000;
 [1]
 RN SEQUENCE FROM N.A.
 RP Fuse S., Tamura M.N.;
 RA "A phylogenetic analysis of the plastid matK gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 CC EMBL; AB040186; BAB16794.1; --
 DR InterPro; IPR000442; Intron_maturase2.
 DR InterPro; IPR002866; MatK_N.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61926 MW; FB689B0255959075 CRC64;
 Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFV 9
 DB 247 FLERIHFV 255
 RESULT 14
 Q9GHC7 PRELIMINARY; PRT; 515 AA.
 AC Q9GHC7;
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Intron maturase (Maturase K).
 GN MATK.
 OS Chamaelirium luteum.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Melanthiaceae;
 OC Chamaelirium.
 OX NCBI_TaxID=112831;
 [1]
 RN SEQUENCE FROM N.A.
 RP Fuse S., Tamura M.N.;
 RA "A phylogenetic analysis of the plastid matK gene with emphasis on
 RT Melanthiaceae sensu lato.";
 RL Plant Biol. 2:415-427(2000).
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 CC EMBL; AB040196; BAB16804.1; --
 DR InterPro; IPR000442; Intron_maturase2.
 DR InterPro; IPR002866; MatK_N.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 KW mRNA processing; Chloroplast.
 SQ SEQUENCE 515 AA; 61365 MW; DC2AC5306073DD09 CRC64;
 Query Match 76.9%; Score 40; DB 8; Length 515;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FLHKVHFV 9
 DB 247 FLERIHFV 255
 RESULT 15
 Q8HV35 PRELIMINARY; PRT; 515 AA.
 AC Q8HV35;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Maturase K.
 GN MATK.
 OS Pyrgophyllum yunnanense.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;
 OC Pyrgophyllum.
 OX NCBI_TaxID=199657;
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Leaf;
 RC Kress W.J., Prince L.M., Williams K.J.;
 RA "The phylogeny and a new classification of the gingers
 RT (Zingiberaceae): Evidence from molecular data.";
 RL Am. J. Bot. 89:1684-1698(2002).
 DR EMBL; AF478878; AAN63235.1; --
 KW Chloroplast.
 SQ SEQUENCE 515 AA; 62132 MW; 99BC93F9EB37635F CRC64;

Query Match 76.9%; Score 40; DB 8; Length 515;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKVHFYV 9

Db 247 FLRIHFYV 255

Search completed: December 16, 2003, 14:17:59
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:05 ; Search time 33 Seconds
(without alignments)
43.289 Million cell updates/sec

Title: US-09-870-089b-7

Perfect score: 61

Sequence: 1 FLHKWHVV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq 19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	100.0	9	23	AAU74683 Human cancer antigen
2	56	91.8	9	23	AAU74684 Human cancer antigen
3	48	78.7	1248	22	ABG19706 Novel human diagnosis
4	42	68.9	172	12	AA10276 Simian immunodeficiency
5	42	68.9	587	19	AAW56104 Euplotes telomeres
6	41	67.2	683	22	ABB59504 Drosophila melanog
7	41	67.2	684	22	ABB57998 Drosophila melanog
8	41	67.2	685	17	ABE89115 Bombyx mori (pro)p
9	41	67.2	685	18	AAW14441 Prophenol oxidase.

10	40	65.6	41	22	AAU74200 Human colon cancer
11	40	65.6	59	21	AB40924 Human ORFX ORF688
12	40	65.6	59	23	ABP31081 Human peptidase-li
13	40	65.6	88	20	AA41219 M. leprae YJ34-MYC
14	40	65.6	123	22	ABG00522 Novel human diagno
15	40	65.6	146	21	ABG15261 Arabidopsis thalia
16	40	65.6	274	22	AB96687 Putative P. abyssal
17	40	65.6	342	23	ABG30742 Gibberella fujikur
18	39	63.9	11	22	AAE12188 Polyglutamine-glut
19	39	63.9	110	23	ABP05237 Human ORFX protein
20	39	63.9	111	22	AAW90060 Human immune/haema
21	39	63.9	166	21	AAAG19315 Arabidopsis thalia
22	39	63.9	166	21	AAAG19315 Arabidopsis thalia
23	39	63.9	210	21	AAAG19314 Arabidopsis thalia
24	39	63.9	210	21	AAAG19314 Arabidopsis thalia
25	39	63.9	217	21	AAAG19313 Arabidopsis thalia
26	39	63.9	217	21	AAAG19313 Arabidopsis thalia
27	39	63.9	360	21	AAAG17199 Arabidopsis thalia
28	39	63.9	529	22	ABBS8340 Drosophila melanog
29	39	63.9	566	21	AAAG17263 Arabidopsis thalia
30	39	63.9	566	21	AAAG17263 Arabidopsis thalia
31	39	63.9	632	21	AAAG17262 Arabidopsis thalia
32	39	63.9	632	21	AAAG17262 Arabidopsis thalia
33	39	63.9	749	22	ABG24802 Novel human diagno
34	38	62.3	79	18	AAW20453 H. pylori secreted
35	38	62.3	79	20	AA41220 H. pylori HEPY pr
36	38	62.3	79	22	AA46322 H. pylori HPS186 p
37	38	62.3	79	22	AA46370 H. pylori HPS186 p
38	38	62.3	80	18	AAW21026 H. pylori secreted
39	38	62.3	98	22	AAAG76780 Human colon cancer
40	38	62.3	158	15	AA363121 ORF1 gene prod. in
41	38	62.3	493	23	ABF38641 Staphylococcus epi
42	38	62.3	548	22	ABH62279 Drosophila melanog
43	38	62.3	556	22	AAU31978 Novel human secret
44	38	62.3	621	22	ABG28304 Novel human diagno
45	38	62.3	1272	22	ABG03108 Novel human diagno

ALIGNMENTS

RESULT 1

ID AAU74683 standard; Peptide; 9 AA.

AC AAU74683;

DT 09-APR-2002 (first entry)

DE Human cancer antigen AFP4/CREB-2 based immunogenic ligand #3.

KW Human; cancer antigen; AFP4; CREB-2; vaccine; cytostatic; immunogenic ligand; gene therapy; MHC; major histocompatibility complex; adoptive immunotherapy; cancer; ovarian cancer.

OS Homo sapiens.

OS Synthetic.

PW WO200192306-A2.

PD 06-DEC-2001.

PF 30-MAY-2001; 2001WO-US17454.

PR 31-MAY-2000; 2000US-209388P.

PR 20-DEC-2000; 2000US-257007P.

PA (GENZ) GENZYME CORP.

PI Nicolette CA;

DR WPI; 2002-097764/13.

XX

PT New therapeutic compounds comprising immunogenic ligands, useful for
PT modulating an immune response, particularly for treating ovarian
PT cancer, and as components of anticancer vaccines -
XX
XX Claim 3; Page 55; 68pp; English.
XX
XX The invention relates to compounds comprising an immunogenic ligand
XX whose sequence is based in part on residues 42-50 of human cancer
XX antigen ATF4/CREB-2 (not defined) and the polynucleotides encoding them.
XX Also included are an antibody that specifically recognises and binds the
XX compound, a method for inducing an immune response in a subject by
XX delivering the compound, a method of immunotherapy comprising
XX administering to a subject the antibody, an immune effector cell that has
XX been raised in vitro or in vivo in the presence and at the expense of an
XX antigen presenting cell that presents the immunogenic compound in the
XX context of an MHC (major histocompatibility complex) molecule and a
XX method of adoptive immunotherapy comprising administering the immune
XX effector cell. The compounds are useful for modulating an immune response
XX to the synthetic and naturally occurring compounds. The compounds are
XX especially useful in gene therapy or as components of anti-cancer
XX vaccines. The compounds are useful for treating cancer, particularly
XX ovarian cancer. The compounds are also useful for generating antibodies
XX that specifically recognise and bind to these molecules. These
XX antibodies are further useful for immunotherapy when administered to a
XX subject. The peptides, polypeptides and polynucleotides are useful in
XX diagnostic methods, for the detection and purification of antibodies,
XX or as immunogens for the production of antibodies. The present
XX sequence represents a human cancer antigen ATF4/CREB-2 based immunogenic
XX ligand of the invention.
XX Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
XX encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
XX but have not been cross-referenced or CDS features put in due to the
XX degeneracy of the DNA sequences.
XX
SQ Sequence 9 AA;
Query Match 100.0%; Score 61; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FLHKWHVVV 9
DB 1 FLHKWHVVV 9
RESULT 2
ID AAU74684 standard; Peptide; 9 AA.
XX
XX AAU74684;
XX
XX 09-APR-2002 (first entry)
XX
XX Human cancer antigen ATF4/CREB-2 based immunogenic ligand #4.
XX
XX Human; cancer antigen; ATF4; CREB-2; vaccine; cytostatic;
XX immunogenic ligand; gene therapy; MHC; major histocompatibility complex;
XX adoptive immunotherapy; cancer; ovarian cancer.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX WO200192306-A2.
XX
XX 06-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US17454.
XX
XX 31-MAY-2000; 2000US-209388P.
XX 20-DEC-2000; 2000US-257007P.
XX
XX (GENZ) GENZYME CORP.
XX

PI Nicolette CA;
XX
XX WPI; 2002-097764/13.
XX
XX New therapeutic compounds comprising immunogenic ligands, useful for
XX modulating an immune response, particularly for treating ovarian
XX cancer, and as components of anticancer vaccines -
XX
XX Claim 4; Page 56; 68pp; English.
XX
XX The invention relates to compounds comprising an immunogenic ligand
XX whose sequence is based in part on residues 42-50 of human cancer
XX antigen ATF4/CREB-2 (not defined) and the polynucleotides encoding them.
XX Also included are an antibody that specifically recognises and binds the
XX compound, a method for inducing an immune response in a subject by
XX delivering the compound, a method of immunotherapy comprising
XX administering to a subject the antibody, an immune effector cell that has
XX been raised in vitro or in vivo in the presence and at the expense of an
XX antigen presenting cell that presents the immunogenic compound in the
XX context of an MHC (major histocompatibility complex) molecule and a
XX method of adoptive immunotherapy comprising administering the immune
XX effector cell. The compounds are useful for modulating an immune response
XX to the synthetic and naturally occurring compounds. The compounds are
XX especially useful in gene therapy or as components of anti-cancer
XX vaccines. The compounds are useful for treating cancer, particularly
XX ovarian cancer. The compounds are also useful for generating antibodies
XX that specifically recognise and bind to these molecules. These
XX antibodies are further useful for immunotherapy when administered to a
XX subject. The peptides, polypeptides and polynucleotides are useful in
XX diagnostic methods, for the detection and purification of antibodies,
XX or as immunogens for the production of antibodies. The present
XX sequence represents a human cancer antigen ATF4/CREB-2 based immunogenic
XX ligand of the invention.
XX Note: Immunogenic ligands AAU74681-AAU74686 are stated to be
XX encoded by the degenerate DNA sequences AAS20120-AAS20125 respectively
XX but have not been cross-referenced or CDS features put in due to the
XX degeneracy of the DNA sequences.
XX
SQ Sequence 9 AA;
Query Match 91.8%; Score 56; DB 23; Length 9;
Best Local Similarity 88.9%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 FLHKWHVVV 9
DB 1 FLHKWHVVV 9
RESULT 3
ID ABG19706 standard; Protein; 1248 AA.
XX
XX ABG19706;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #19697.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS83893.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID No 50065; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1248 AA;
 Query Match 78.7%; Score: 48; DB 22; Length 1248;
 Best Local Similarity 75.0%; Pred. No. 78;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LHKWHWV 9
 Db 1065 LHKWHWV 1072
 RESULT 4
 AAR10276
 ID AAR10276 standard; Protein; 172 AA.
 AC AAR10276;
 XX
 XX 04-APR-1991 (first entry)
 DT
 XX Simian immunodeficiency virus vif protein.
 XX SIVmnd; HIV; AIDS; vaccine; pSMH103.
 XX
 XX Simian immunodeficiency virus.
 XX JP02299587-A.
 XX
 XX 11-DEC-1990.
 PD
 XX 11-MAY-1989; 89JP-0116129.
 PF
 XX 11-MAY-1989; 89JP-0116129.
 PR
 XX (TOFU) TONEN CORP.
 PA
 XX WPI; 1991-027566/04.
 DR
 XX N-PSDB; AAQ10203.
 DR

XX New complementary DNA to RNA gene of simian immuno-deficiency
 PT virus - used for preparing vaccine and diagnosis agent for AIDS
 XX
 XX Claim 2; Fig 4; 14pp; Japanese.
 XX
 CC SIV proteins may be produced from an E.coli expression system
 CC transformed with plasmid pSMH103.
 CC These may be used in diagnosis, treatment, and development of a vaccine
 CC against AIDS, as the sequence has the same structure as HIV-1 (but
 CC lacking the vpx gene).
 XX
 SQ Sequence 172 AA;
 Query Match 68.9%; Score 42; DB 12; Length 172;
 Best Local Similarity 62.5%; Pred. No. 84;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LHKWHWV 9
 Db 19 LHKWHWV 26
 RESULT 5
 AAW56104
 ID AAW56104 standard; Protein; 587 AA.
 XX AAW56104;
 AC
 XX 13-AUG-1998 (first entry)
 DT
 XX Euplotes telomerase reverse transcriptase subunit 1st reading frame.
 DE Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;
 XX prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.
 KW
 XX Euplotes sp.
 OS
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 11 /label= stop codon
 FT /note= "encoded by TAA"
 FT Misc-difference 18 /label= stop codon
 FT /note= "encoded by TAG"
 FT Misc-difference 38 /label= stop codon
 FT /note= "encoded by TAG"
 FT Misc-difference 62 /label= stop codon
 FT /note= "encoded by TGA"
 FT Misc-difference 71 /label= stop codon
 FT /note= "encoded by TAG"
 FT Misc-difference 80 /label= stop codon
 FT /note= "encoded by TGA"
 FT Misc-difference 95 /label= stop codon
 FT /note= "encoded by TAG"
 FT Misc-difference 98 /label= stop codon
 FT /note= "encoded by TAG"
 FT Misc-difference 100 /label= stop codon
 FT /note= "encoded by TAA"
 FT Misc-difference 107 /label= stop codon
 FT /note= "encoded by TAG"
 FT Misc-difference 113 /label= stop codon
 FT /note= "encoded by TAA"
 FT Misc-difference 120

FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 121
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 140
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 142
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 155
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 156
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 159
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 187
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 188
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 213
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 218
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 221
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 233
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 236
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 245
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 281
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 298
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 359
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 380
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 398
FT /label= stop codon
FT /note= "encoded by TAG"
FT Misc-difference 406
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 410
FT /label= stop codon
FT /note= "encoded by TAG"
FT Misc-difference 415
FT /label= stop codon
FT /note= "encoded by TAG"
FT Misc-difference 430
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 431
FT /label= stop codon

FT /note= "encoded by TAA"
FT Misc-difference 439
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 441
FT /label= stop codon
FT /note= "encoded by TAG"
FT Misc-difference 446
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 472
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 476
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 477
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 488
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 497
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 506
FT /label= stop codon
FT /note= "encoded by TAG"
FT Misc-difference 510
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 524
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 528
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 530
FT /label= stop codon
FT /note= "encoded by TAG"
FT Misc-difference 557
FT /label= stop codon
FT /note= "encoded by TAA"
FT Misc-difference 560
FT /label= stop codon
FT /note= "encoded by TGA"
FT Misc-difference 564
FT /label= stop codon
FT /note= "encoded by TGA"
FT XX
FT GB2317891-A,
FT XX
FT 08-APR-1998.
FT XX
FT 01-OCT-1997; 97GB-0020890.
FT XX
FT 14-AUG-1997; 97US-0915503.
FT 01-OCT-1996; 96US-0724543.
FT 18-APR-1997; 97US-0844419.
FT 25-APR-1997; 97US-0846017.
FT 06-MAY-1997; 97US-0851843.
FT 09-MAY-1997; 97US-0854050.
FT 14-AUG-1997; 97US-0911312.
FT 14-AUG-1997; 97US-0912951.
FT XX
FT (GERO-) GERON CORP.
FT (UYTE-) UNIV TECHNOLOGY CORP.
FT XX
FT Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J;
FT Morin GB, Nakamura T, Harley CB;
FT WPI; 1998-171633/16.
FT N-PSDB; AAV22422.
FT DR

Pure and recombinant human Telomerase Reverse Transcriptase and its variants - are useful in the diagnosis, prognosis and treatment of cell proliferation conditions especially cancer and ageing
Example 1; Fig 35; 387pp; English.
Query Match 68.9%; Score 42; DB 19; Length 587;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 1 FLHKHWVV 9 :: Dd 25 FPKHWEIL 33
RESULT 6
ABB59504
ID ABB59504 standard; Protein; 683 AA.
XX AC ABB59504;
XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster polypeptide SEQ ID NO 5304.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
PD 23-MAR-2001; 2001WO-USO9231.
PF 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABLO3607.
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT Disclosure; SEQ ID NO 5304; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA sequences (ABLI01840-ABLI16175) and the encoded proteins (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 683 AA;
SQ Query Match 67.2%; Score 41; DB 22; Length 683;
Best Local Similarity 83.3%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 LHKWHW 7

Tue Dec 16 15:00:22 2003

KW Silkworm; larvae; pro; phenoloxidase; Phe52-Val693; Phe52-Gly685
KW labelling oxidase.

XX Bombyx mori.

XX Location/Qualifiers

XX Key

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

XX Peptide

XX /note= "peptide fragment"

PF 13-JUL-1995; 95JP-0177444.

XX 13-JUL-1995; 95JP-0177444.

XX (WAKP) WAKO PURE CHEM IND LTD.

XX WPI; 1997-148588/14.

XX N-PSDB; AAT62660.

XX (Pro)phenol oxidase derived from a domestic silkworm - useful as a

XX labelling oxidase and in pro-phenol oxidase activation system for

XX detection of microorganisms

XX Claim 2; Page 14-17; 18pp; Japanese.

XX This protein is a prophenol oxidase derived from a domestic silkworm.

XX The phenol oxidase may be used as a novel labelling oxidase. The

XX elucidation of the primary structure of the prophenol oxidase will

XX contribute to the reconstitution of a prophenol oxidase activation system

XX which can be applied to the detection of microorganisms by measurement of

XX beta-1,3-glucan and peptide glycan.

XX Sequence 685 AA;

XX Query Match 67.2%; Score 41; DB 18; Length 685;

XX Best Local Similarity 83.3%; Pred. No. 4.3e+02;

XX Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

XX QY 2 LHKWHW 7

XX DB 207 LHHWHW 212

XX RESULT 10

XX AAG74200

XX ID AAG74200 standard; Protein; 41 AA.

XX AC AAG74200;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen protein SEQ ID NO:4964.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX KW colorectal carcinoma.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX DR WPI; 2001-235357/24.

XX DR N-PSDB; AAH33631.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 11; Page 6720; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

XX cancer-associated nucleic acid molecules (N) and proteins (P), where

XX the proteins are collectively known as colon cancer antigens. The colon

XX cancer antigens have cytostatic activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAB37196 to AAB37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 41 AA;

Query Match 65.6%; Score 40; DB 22; Length 41;
 Best Local Similarity 50.0%; Pred. No. 42;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWV 9
 |||:::
 Db 27 LHPWNWII 34

RESULT 11

AAB40924

ID AAB40924 standard; Protein; 59 AA.

XX AAB40924;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF688 polypeptide sequence SEQ ID NO:1376.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinketsu RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAB75133.

XX Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 11; Page 1172; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus, lupus
 CC erythematosus, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 59 AA;

Query Match 65.6%; Score 40; DB 21; Length 59;
 Best Local Similarity 83.3%; Pred. No. 59;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHW 7
 |||:::
 Db 19 LHKWHW 24

RESULT 12

ABP31081

ID ABP31081 standard; Protein; 59 AA.

XX ABP31081;

XX 09-JUL-2002 (first entry)

XX Human peptidase-like ORF54 protein, SEQ ID NO:108.

KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibition; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnary;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide.

XX Homo sapiens.

XX WO200190366-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US17076.

XX 24-MAY-2000; 2000US-206690P.

XX (CURA-) CURAGEN CORP.

XX Leach MD, Shimkets RA;
 XX WPI: 2002-106200/14.
 DR N-PSDB; ABN75107.
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation -
 XX Claim 10; Page 297; 2508pp; English.
 PS Sequences ABP31028-ABP3561 represent 4534 novel human proteins
 XX designated ORF (Open reading frame) 1-4534, and sequences ABN75054-
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
 CC and antiinfective activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,
 CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes, in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.
 XX
 SQ Sequence 59 AA;
 Query Match 65.6%; Score 40; DB 23; Length 59;
 Best Local Similarity 83.3%; Pred. No. 59;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LHKWHW 7
 |||||
 Db 19 LRGWHEW 24
 RESULT 13
 AAY41219
 ID AAY41219 standard; Protein; 88 AA.
 XX
 AC AAY41219;
 XX
 DT 31-JAN-2000 (first entry)
 XX
 DE M. leprae YJ34-WYCLE protein.
 XX
 KW Membrane targeting protein; translocation protein; Escherichia coli;
 KW Sec-independent pathway; protein transport; twin-arginine signal peptide;
 KW mtABC operon; MtaA protein.
 XX

OS Mycobacterium leprae.
 XX WO9951753-A1.
 XX PD 14-OCT-1999.
 XX PF 29-MAR-1999; 99WO-CA00272.
 XX 01-APR-1998; 98US-0053197.
 PR 28-MAY-1998; 98US-0085761.
 XX (UYAL-) UNIV ALBERTA.
 XX Weiner JH, Turner RJ;
 XX WPI: 1999-633740/54.
 DR
 XX New recombinant membrane targeting and translocation proteins from
 PT Escherichia coli, used to produce soluble polypeptides normally
 PT produced in insoluble form -
 XX Example 6; Fig 8; 11pp; English.
 PS The invention relates to recombinant membrane targeting and
 XX translocation proteins from Escherichia coli. The recombinant
 CC polypeptides are involved in the Sec-independent pathway for transporting
 CC proteins with a twin- arginine signal peptide to the periplasm or
 CC extracellular medium and to the cell membrane. They transport fully
 CC folded and co-factor containing proteins. The polypeptides are used to
 CC transport such proteins, specifically to produce soluble forms of
 CC polypeptides that are normally produced in insoluble form. They may also
 CC be used to raise specific antibodies. Nucleic acid sequences that encode
 CC the polypeptides are used for production of recombinant proteins and
 CC or cDNA sequences (these have been found in many other bacteria, yeast,
 CC plants, nematodes and humans). Producing normally insoluble proteins in
 CC soluble form eliminates the need for resolubilization with strong
 CC denaturants and facilitates recovery of functional proteins (which have
 CC been properly folded by cytoplasmic enzymes before translocation).
 CC Sequences AAY41214-41223 represent proteins homologous to the E. coli
 CC MtaA protein.
 XX
 SQ Sequence 88 AA;
 Query Match 65.6%; Score 40; DB 20; Length 88;
 Best Local Similarity 75.0%; Pred. No. 86;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 LHKWHWVWV 9
 | | | | |
 Db 4 LSPWHWVWV 11
 RESULT 14
 ABG00522
 ID ABG00522 standard; Protein; 123 AA.
 XX
 AC ABG00522;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #513.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 FN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS64709.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID NO 30881; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 123 AA;
Query Match 65.6%; Score 40; DB 22; Length 123;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 2 LHKWHWV 8
Db 44 LHKWHWV 50
RESULT 15
AAG15261
ID AAG15261 standard; Protein; 146 AA.
XX
XX
XX AAG15261;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 15445.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140153.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 65.6%; Score 40; DB 21; Length 146;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLHKWHWV 8
Db 64 FLLTWHWV 71

Search completed: December 16, 2003, 14:14:28
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:11:00 ; Search time 12.3333 Seconds
(without alignments)
30.875 Million cell updates/sec

Title: US-09-870-089B-7
Sequence: 1 FLHKWHWV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pcp:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pcp:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pcp:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pcp:*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pcp:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	42	68.9	19	US-08-974-549A-118	Sequence 118, Appl
2	42	68.9	552	US-08-851-843A-4	Sequence 4, Appl
3	42	68.9	552	US-08-854-050-4	Sequence 4, Appl
4	42	68.9	552	US-09-430-323-4	Sequence 4, Appl
5	40	65.6	88	US-09-053-197A-14	Sequence 14, Appl
6	40	65.6	88	US-09-085-761A-14	Sequence 14, Appl
7	39	63.9	104	US-09-060-726A-6	Sequence 6, Appl
8	39	63.9	143	US-09-732-210-615	Sequence 615, Appl
9	38	62.3	79	US-09-053-197A-15	Sequence 15, Appl
10	38	62.3	79	US-09-085-761A-15	Sequence 15, Appl
11	38	62.3	158	US-08-411-768B-5	Sequence 5, Appl
12	38	62.3	493	US-09-134-001C-3486	Sequence 3486, Ap
13	37	60.7	191	US-09-107-532A-5382	Sequence 5382, Ap
14	37	60.7	246	US-09-199-637A-407	Sequence 407, Appl
15	37	60.7	442	US-09-328-352-5670	Sequence 5670, Ap
16	36	59.0	15	US-08-300-386A-35	Sequence 35, Appl
17	36	59.0	15	US-08-931-645-35	Sequence 35, Appl
18	36	59.0	15	PCT-US94-01258-35	Sequence 35, Appl
19	36	59.0	15	PCT-US95-11235-35	Sequence 35, Appl
20	36	59.0	118	US-09-343-698-6	Sequence 6, Appl
21	36	59.0	130	US-08-451-947-67	Sequence 67, Appl
22	36	59.0	130	US-08-424-826A-67	Sequence 67, Appl
23	36	59.0	130	US-08-928-694-67	Sequence 67, Appl
24	36	59.0	130	US-08-450-842-67	Sequence 67, Appl
25	36	59.0	130	US-08-451-390-67	Sequence 67, Appl
26	36	59.0	130	PCT-US91-06950-67	Sequence 67, Appl
27	36	59.0	343	US-09-413-814-99	Sequence 99, Appl

US-08-974-549A-118
; Sequence 118, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951

ALIGNMENTS

RESULT 1

US-08-974-549A-118
; Sequence 118, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951

Sequence 28574, A
Sequence 2, Appli
Sequence 7, Appli
Sequence 8, Appli
Sequence 9, Appli
Sequence 116, Appl
Sequence 13, Appli
Sequence 7, Appli
Sequence 4203, Ap
Sequence 31177, A
Sequence 4, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 4, Appli
Sequence 5, Appli
Sequence 34, Appli

; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA: US 08/915,503
; APPLICATION NUMBER: 01-OCT-1997
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-974-549A-118

Query Match 68.9%; Score 42; DB 3; Length 19;
Best Local Similarity 55.6%; Pred. No. 3.2;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLHKRWVV 9
| | | | | : :
Db 7 FPHKRWIL 15

RESULT 2
US-08-851-843A-4
; Sequence 4, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851.843A
; FILING DATE: 06-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-851-843A-4

Query Match 68.9%; Score 42; DB 3; Length 552;
Best Local Similarity 55.6%; Pred. No. 63;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLHKRWVV 9
| | | | | : :
Db 23 FPHKRWIL 31

RESULT 3
US-08-854-050-4
; Sequence 4, Application US/08854050
; Patent No. 6261836
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6261836el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; CLASSIFICATION:

RESULT 6

Query Match 63.9%; Score 39; DB 3; Length 104;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0

ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Review #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/053,197A
FILING DATE: 01-APR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: UALB-03293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8938

; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-053-197A-15

Query Match 62.3%; Score 38; DB 3; Length 79;
Best Local Similarity 55.6%; Pred. No. 42;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FLHKWHVV 9
DB 4 FTSIWHVI 12

RESULT 10
US-09-085-761A-15
; Sequence 15, Application US/09085761A
; Patent No. 6335178
; GENERAL INFORMATION:
; APPLICANT: Weiner, Joel H.
; APPLICANT: Turner, Raymond J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN
; TITLE OF INVENTION: SECRETION
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,761A
; FILING DATE: 28-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: UALB-03356
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-085-761A-15

Query Match 62.3%; Score 38; DB 4; Length 79;
Best Local Similarity 55.8%; Pred. No. 42;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FLHKWHVV 9
DB 4 FTSIWHVI 12

RESULT 11
US-08-411-768B-5
; Sequence 5, Application US/08411768B
; Patent No. 6083712

; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-411-768B-5

Query Match 62.3%; Score 38; DB 3; Length 158;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 WHWV 9
DB 62 WHWV 66

RESULT 12
US-09-134-001C-3486
; Sequence 3486, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3486
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3486

Query Match 62.3%; Score 38; DB 4; Length 493;
Best Local Similarity 57.1%; Pred. No. 2.1e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWV 8
Db 167 LDRWHWI 173

RESULT 13
US-09-107-532A-5382
; Sequence 5382, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5382:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...191
; SEQUENCE DESCRIPTION: SEQ ID NO: 5382:
US-09-107-532A-5382

Query Match 60.7%; Score 37; DB 4; Length 191;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWV 9
Db 52 LHQYHWV 59

RESULT 14
US-09-199-637A-407
; Sequence 407, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.

Query Match 60.7%; Score 37; DB 4; Length 246;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 5 WHWV 9
Db 235 WHWV 239

RESULT 15
US-09-328-352-5670
; Sequence 5670, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Brston et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5670
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5670

Query Match 60.7%; Score 37; DB 4; Length 442;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 FLHKWHWV 9
Db 222 FVFPWHWV 230

Search completed: December 16, 2003, 14:20:40
Job time : 13.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:18:11 ; Search time 22.1667 Seconds
(without alignments)
75.512 Million cell updates/sec

Title: US-09-870-089B-7
Perfect score: 61
Sequence: 1 FLHKWHVV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	100.0	9	12	US-09-870-089B-7
2	56	91.8	9	12	US-09-870-089B-9
3	42	68.9	552	10	US-09-843-676-4
4	42	68.9	552	10	US-09-766-253-4
5	42	68.9	552	11	US-09-438-486-4
6	42	68.9	552	15	US-10-053-758-4
7	42	68.9	552	15	US-10-054-295-4
8	42	68.9	552	15	US-10-054-611-4
9	40	65.6	41	15	US-10-106-698-4974
10	39	63.9	11	9	US-09-780-070-21
11	39	63.9	106	9	US-09-291-809C-8
12	39	63.9	106	9	US-09-845-849-8
13	38	62.3	98	15	US-10-106-698-7554
14	37	60.7	9	12	US-09-870-089B-3
15	37	60.7	19	12	US-10-029-386-33052

16	37	60.7	246	11	US-09-975-719-407	Sequence 407, Appl
17	37	60.7	831	12	US-10-100-294A-3	Sequence 3, Appli
18	36	59.0	115	11	US-09-798-889-56	Sequence 56, Appl
19	36	59.0	130	8	US-08-450-842-67	Sequence 67, Appl
20	36	59.0	412	15	US-10-156-761-9146	Sequence 9146, Ap
21	36	59.0	501	15	US-10-156-761-12720	Sequence 12720, A
22	36	59.0	555	12	US-10-137-870-72	Sequence 72, Appl
23	36	59.0	555	12	US-10-140-018-72	Sequence 72, Appl
24	36	59.0	555	12	US-10-140-021-72	Sequence 72, Appl
25	36	59.0	555	12	US-10-140-274-72	Sequence 72, Appl
26	36	59.0	555	12	US-10-140-471-72	Sequence 72, Appl
27	36	59.0	555	12	US-10-140-807-72	Sequence 72, Appl
28	36	59.0	555	12	US-10-140-922-72	Sequence 72, Appl
29	36	59.0	555	12	US-10-140-924-72	Sequence 72, Appl
30	36	59.0	555	12	US-10-140-926-72	Sequence 72, Appl
31	36	59.0	555	12	US-10-141-698-72	Sequence 72, Appl
32	36	59.0	555	12	US-10-141-702-72	Sequence 72, Appl
33	36	59.0	555	12	US-10-141-704-72	Sequence 72, Appl
34	36	59.0	555	12	US-10-142-421-72	Sequence 72, Appl
35	36	59.0	555	12	US-10-142-432-72	Sequence 72, Appl
36	36	59.0	555	12	US-10-142-767-72	Sequence 72, Appl
37	36	59.0	555	12	US-10-143-033-72	Sequence 72, Appl
38	36	59.0	555	12	US-10-144-994-72	Sequence 72, Appl
39	36	59.0	555	12	US-10-145-628-72	Sequence 72, Appl
40	36	59.0	555	12	US-10-145-631-72	Sequence 72, Appl
41	36	59.0	555	12	US-10-145-633-72	Sequence 72, Appl
42	36	59.0	555	12	US-10-145-746-72	Sequence 72, Appl
43	36	59.0	555	12	US-10-145-748-72	Sequence 72, Appl
44	36	59.0	555	12	US-10-145-823-72	Sequence 72, Appl
45	36	59.0	555	12	US-10-145-826-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1

US-09-870-089B-7
; Sequence 7, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATP4/CREB-2
US-09-870-089B-7

Query Match 100.0%; Score 61; DB 12; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLHKWHVV 9

Db 1 FLHKWHVV 9

RESULT 2

US-09-870-089B-9
; Sequence 9, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870,089B

;
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATP4/CREB-2
US-09-870-089B-9

Query Match 91.8%; Score 56; DB 12; Length 9;
Best Local Similarity 88.9%; Pred. No. 6.1e+05;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLHKWHWV 9
| | | | | | | | | |
Db 1 FLHKWHWV 9

RESULT 3
US-09-843-676-4
; Sequence 4, Application US/09843676
; Patent No. US20020164786A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin
; Andrews, William H.
; TITLE OF INVENTION: No. US20020164786A1el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/843,676
; FILING DATE: 26-Apr-2001
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. US20020164786A1 Relevant
; TOPOLOGY: No. US20020164786A1 Relevant
; MOLECULE TYPE: protein

;
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-843-676-4
Query Match 68.9%; Score 42; DB 10; Length 552;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 FLHKWHWV 9
| | | | | | | | | |
Db 23 FPHKWRWIL 31
RESULT 4
US-09-766-253-4
; Sequence 4, Application US/09766253
; Publication No. US20020187471A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin
; Andrews, William H.
; TITLE OF INVENTION: No. US20020187471A1el Telomerase
; NUMBER OF SEQUENCES: 171
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/766,253
; FILING DATE: 19-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/846,017
; FILING DATE: 1997-04-25
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002920US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-766-253-4
Query Match 68.9%; Score 42; DB 10; Length 552;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 FLHKWHWV 9
| | | | | | | | | |
Db 23 FPHKWRWIL 31

RESULT 5
US-09-438-486-4
; Sequence 4, Application US/09438486
; Publication No. US20030009019A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. US20030009019A1el Telomerase
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 12-NOV-1999
; APPLICATION NUMBER: US/09/438,486
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002931US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-09-438-486-4

Query Match 68.9%; Score 42; DB 11; Length 552;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLKHQHVWV 9
DB 23 FPHKRWIL 31

RESULT 6
US-10-053-758-4

; Sequence 4, Application US/10053758
; Publication No. US20030032075A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. US20030032075A1el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/053,758
; FILING DATE: 18-Jan-2002
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. US20030032075A1 Relevant
; TOPOLOGY: No. US20030032075A1 Relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-053-758-4

Query Match 68.9%; Score 42; DB 15; Length 552;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLKHQHVWV 9
DB 23 FPHKRWIL 31

RESULT 7
US-10-054-295-4
; Sequence 4, Application US/10054295
; Publication No. US20030044953A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru

Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin
Andrews, William H.
TITLE OF INVENTION: No. US20030044953A1el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/054,295
FILING DATE: 18-Jan-2002
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/854,050
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 552 amino acids
TYPE: amino acid
STRANDEDNESS: No. US20030044953A1 Relevant
TOPOLOGY: No. US20030044953A1 Relevant
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-054-295-4
Query Match 68.9%; Score 42; DB 15; Length 552;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 FLHKWHWV 9
Db 23 FPHKRWIL 31
RESULT 8
US-10-054-611-4
Sequence 4, Application US/10054611
Publication No. US20030059787A1
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin
Andrews, William H.
TITLE OF INVENTION: No. US20030059787A1el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/054,611
FILING DATE: 18-Jan-2002
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/854,050
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 552 amino acids
TYPE: amino acid
STRANDEDNESS: No. US20030059787A1 Relevant
TOPOLOGY: No. US20030059787A1 Relevant
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-054-611-4
Query Match 68.9%; Score 42; DB 15; Length 552;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 FLHKWHWV 9
Db 23 FPHKRWIL 31
RESULT 9
US-10-106-698-4974
Sequence 4974, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 4974
LENGTH: 41
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE

; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-4974

Query Match 65.6%; Score 40; DB 15; Length 41;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHVV 9
Db 27 LHPWNWII 34

RESULT 10
US-09-780-070-21
; Sequence 21, Application US/09780070
; Patent No. US2002009752A1
; GENERAL INFORMATION:
; APPLICANT: Burke, James
; APPLICANT: Strittmatter, Warren
; APPLICANT: Nagai, Yoshitaka
; TITLE OF INVENTION: COMPOUNDS THAT SELECTIVELY BIND TO EXPANDED POLYGLUTAMINE REPEAT
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: 5405.242
; CURRENT APPLICATION NUMBER: US/09/780.070
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/189,781
; PRIOR FILING DATE: 2000-03-16
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 11
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-780-070-21

Query Match 63.9%; Score 39; DB 9; Length 11;
Best Local Similarity 80.0%; Pred. No. 29;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HKWHW 7
Db 1 HENHW 5

RESULT 11
US-09-291-809C-8
; Sequence 8, Application US/09291809C
; Patent No. US2001004981A1
; GENERAL INFORMATION:
; APPLICANT: Detlef Weigel
; APPLICANT: Salk Institute
; TITLE OF INVENTION: FLOWERING LOCUS T (FT) AND GENETICALLY
; TITLE OF INVENTION: MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
; FILE REFERENCE: SALKINS.026CP1
; CURRENT APPLICATION NUMBER: US/09/291,809C
; CURRENT FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: PCT/US99/08151
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 09/060,726
; PRIOR FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-291-809C-8

Query Match 63.9%; Score 39; DB 9; Length 106;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KWHVV 9
Db 50 EHWVV 55

RESULT 12
US-09-845-849-8
; Sequence 8, Application US/09845849
; Patent No. US20020029395A1
; GENERAL INFORMATION:
; APPLICANT: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
; APPLICANT: WEIGEL, Detlef
; APPLICANT: KARDAILSKY, Igor
; TITLE OF INVENTION: FLOWERING LOCUS T (FT) AND GENETICALLY
; TITLE OF INVENTION: MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
; FILE REFERENCE: SALKINS.026DV1
; CURRENT APPLICATION NUMBER: US/09/845,849
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/060,726
; PRIOR FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-845-849-8

Query Match 63.9%; Score 39; DB 9; Length 106;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KWHVV 9
Db 50 EHWVV 55

RESULT 13
US-10-106-698-7554
; Sequence 7554, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; *PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 7554
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (65)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (85)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (88)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (98)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7554

Search completed: December 16, 2003, 14:41:06
Job time : 23.1667 secs

Query Match 62.3%; Score 38; DB 15; Length 98;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLKHWHVV 9
||:|:|
Db 5 FLQWFWLV 13

RESULT 14
US-09-870-089B-3
; Sequence 3, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881203900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATP4/CREB-2
US-09-870-089B-3

Query Match 60.7%; Score 37; DB 12; Length 9;
Best Local Similarity 66.7%; Pred. No. 6.1e+05;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLKHWHVV 9
||:|:|
Db 1 FLXWHGFV 9

RESULT 15
US-10-029-386-33052
; Sequence 33052, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR Q
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33052
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z49918.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.91
US-10-029-386-33052

Query Match 60.7%; Score 37; DB 12; Length 19;
Best Local Similarity 66.7%; Pred. No. 84;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKWHW 7
:|:|:|
Db 4 VNKWHW 9

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:10:15 ; Search time 10.3333 Seconds
(without alignments)
83.760 Million cell updates/sec

Title: US-09-870-089B-7
Perfect score: 61
Sequence: 1 FLHKWHVV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: PIR1.*
- 2: PIR2.*
- 3: PIR3.*
- 4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	75.4	209	2 A87628	conserved hypothet
2	46	75.4	218	2 G82786	conserved hypothet
3	42	68.9	172	2 S28082	vif protein - simi
4	41	67.2	582	2 F95871	hypothetical prote
5	41	67.2	609	2 G84832	ATP-dependent RNA
6	40	65.6	88	1 S72850	hypothetical prote
7	40	65.6	95	2 E87075	probable secreted
8	40	65.6	260	2 G75174	hypothetical prote
9	40	65.6	295	2 H71039	hypothetical prote
10	40	65.6	1663	2 T28923	hypothetical prote
11	39	63.9	143	2 E64359	ribosomal protein
12	39	63.9	145	2 A70014	hypothetical prote
13	39	63.9	159	2 T75311	conserved hypothet
14	39	63.9	210	2 T06656	hypothetical prote
15	39	63.9	259	2 AG2462	hypothetical prote
16	39	63.9	795	2 T20609	hypothetical prote
17	39	63.9	1059	2 T21891	hypothetical prote
18	39	63.9	3085	2 T00327	polyprotein - infe
19	38	62.3	79	1 H64559	conserved hypothet
20	38	62.3	79	2 A71948	hypothetical prote
21	38	62.3	158	1 Q3ECBA	conserved hypothet
22	38	62.3	158	2 C90735	hypothetical prote
23	38	62.3	158	2 D85585	hypothetical prote
24	38	62.3	158	2 AF0586	conserved hypothet
25	38	62.3	179	2 T36855	conserved hypothet
26	38	62.3	193	2 JN0846	enterohemolysin 2
27	38	62.3	257	2 A99775	hypothetical prote
28	38	62.3	257	2 D85637	hypothetical prote
29	38	62.3	317	2 AF2129	hypothetical prote

ALIGNMENTS

RESULT 1

A87628

conserved hypothetical protein CC3059 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C;Accession: A87628

R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kol

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: A87628

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-209 <STO>

A;Cross-references: GB:AE005673; NID:gl3424707; PIDN:AAK25021.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC3059

C;Superfamily: Defnococcus radiodurans hypothetical protein DR1792

Query Match	75.4%	Score 46;	DB 2;	Length 209;
Best Local Similarity	71.4%	Pred. No. 3.9;		
Matches	5;	Conservative	2;	Mismatches 0; Indels 0; Gaps 0;
Q	2 LHKWHVV 8			
Db	17 LHQWHVI 23			

RESULT 2

G82786

conserved hypothetical protein XF0597 [imported] - Xylella fastidiosa (strain 9a5c)

C;Species: Xylella fastidiosa

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C;Accession: G82786

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: G82786

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-218 <SIM>

A;Cross-references: GB:AE003905; GB:AE003849; NID:g9105456; PIDN:AAF83407.1; GSPDB:GN0

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Rainach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier,

as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froi

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshahako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XP0597
C;Superfamily: Deinococcus radiodurans hypothetical protein DR1792

Query Match 75.4%; Score 46; DB 2; Length 218;
Best Local Similarity 71.4%; Pred. No. 4;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKWHWV 8
||:||||:
Db 28 LKQWHWI 34

RESULT 3
S28082
vif protein - simian immunodeficiency virus
C;Species: simian immunodeficiency virus, SIV
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Nov-1999
C;Accession: S28082
R;Tsujimoto, H.; Hasegawa, A.; Maki, N.; Fukasawa, M.; Miura, T.; Speidel, S.; Cooper, R
Nature 341, 539-541, 1989
A;Title: Sequence of a novel simian immunodeficiency virus from a wild-caught African ma
A;Reference number: S28080; MUID:90015168; PMID:2797181
C;Accession: S28082
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-172 <TSU>
A;Cross-references: GB:M27470; EMBL:X15781; NID:G334583; PIDN:AAB49570.1; PID:G334588
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1989
C;Genetics:
A;Gene: vif
C;Superfamily: AIDS vif protein

Query Match 68.9%; Score 42; DB 2; Length 172;
Best Local Similarity 62.5%; Pred. No. 13;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWV 9
:||||:
Db 19 IEKWHWLV 26

RESULT 4
F95871
hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) megaplasamid pSymB
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: F95871
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernar
proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1.683-kb pSymB megaplasamid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: F95871
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-582 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC48638.1; PID:g15140110; GSPDB:GN00167
A;Experimental source: Strain 1021, megaplasamid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: Smb20248
A;Genome: plasmid

Query Match 67.2%; Score 41; DB 2; Length 582;
Best Local Similarity 62.5%; Pred. No. 61;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWV 9
||:||||:
Db 463 LHRWEWLV 470

RESULT 5

G84832
ATP-dependent RNA helicase [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84832

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: G84832

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-609 <STO>

A;Cross-references: GB:AE002093; NID:G4895231; PIDN:AAD32817.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g40700

A;Map position: 2

Query Match 67.2%; Score 41; DB 2; Length 609;
Best Local Similarity 62.5%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKWHWV 9
||:||||:
Db 128 LHRFWIV 135

RESULT 6

S72850
hypothetical protein B2126_C1_182 - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 23-Mar-2001

C;Accession: S72850

R;Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A;Description: Mycobacterium leprae cosmid B2126.

A;Reference number: S72585

A;Accession: S72850

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-88 <SMI>

A;Cross-references: EMBL:U00017; NID:G4666994; PID:G467005

C;Genetics:

A;Start codon: GTG

C;Superfamily: conserved hypothetical secreted protein HP0320

Query Match 65.6%; Score 40; DB 1; Length 88;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LHKWHWV 9
||:||||:
Db 4 LSPWHWV 11

RESULT 7
E87075
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 17-May-2002
C;Accession: E87075
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sg
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: E87075
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-95 <STO>
A;Cross-references: GB:AL450380; NID:gl3093243; PIDN:CAC31712.1; GSPDB:GN00147
C;Genetics:
C;Gene: MLJ331
C;Superfamily: conserved hypothetical secreted protein HP0320

Query Match 65.6%; Score 40; DB 2; Length 95;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LHKWHVV 9
| | | | |
Db 11 LSPWHVV 18

RESULT 8
G75174
hypothetical protein PAB1992 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: G75174
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: G75174
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-260 <KAW>
A;Cross-references: GB:AJ248284; GB:AL096836; NID:G5457730; PIDN:CAB49478.1; PID:G545798
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1992

Query Match 65.6%; Score 40; DB 2; Length 260;
Best Local Similarity 71.4%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLHKWHV 7
| | | | |
Db 52 FFHKWSW 58

RESULT 9
H71039
hypothetical protein PH1608 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: H71039
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: H71039
A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA
A;Residues: 1-295 <KAW>
A;Cross-references: GB:AF000006; NID:g3236133; PIDN:BAA30720.1; PID:g3258037
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1608

Query Match 65.6%; Score 40; DB 2; Length 295;
Best Local Similarity 71.4%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLHKWHV 7
| | | | |
Db 86 FFHKWSW 92

RESULT 10
T28923
hypothetical protein C48B6.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28923
R;Minx, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid C48B6.
A;Reference number: Z20544
A;Accession: T28923
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1663 <MIN>
A;Cross-references: EMBL:U97189; PIDN:AAC48166.1; GSPDB:GN00019; CESP:C48B6.6
A;Experimental source: strain Bristol N2; clone C48B6
C;Genetics:
A;Gene: CESP:C48B6.6
A;Map position: 1
A;Introns: 43/2; 111/1; 185/3; 221/3; 298/2; 327/3; 364/2; 411/2; 500/1; 595/1; 643/3; 2;
2; 1594/1; 1638/3

Query Match 65.6%; Score 40; DB 2; Length 1663;
Best Local Similarity 44.4%; Pred. No. 2.5e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FLHKWHVV 9
| | | | |
Db 75 FIEKWDVII 83

RESULT 11
B64359
ribosomal protein L15 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: B64359
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodok, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: B64359
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-143 <BUL>
A;Cross-references: GB:U67497; GB:L77117; NID:G2826284; PIDN:AAB98466.1; PID:gl591179; T
C;Genetics:
A;Map position: FOR419631-420062
C;Superfamily: rat ribosomal protein L27a

Query Match 63.9%; Score 39; DB 2; Length 143;
Best Local Similarity 57.1%; Pred. No. 31;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

3 HKWHWVV 9
|||
42 HKWTWII 48

SULT 12

0014

potheical protein yuku - Bacillus subtilis

Species: Bacillus subtilis

Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

Accession: A70014

Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, J.; Carter, N.M.; Ch
; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
ture 390, 249-256, 1997

Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
sch, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
etter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, E.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
; Authors: Schleich, S.; Schroter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
; Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
; Reference number: A69580; MUID:98044033; PMID:9384377

Accession: A70014

Status: preliminary; nucleic acid sequence not shown; translation not shown

Molecule type: DNA

Residues: 1-145 <KUN>

Cross-references: GB:Z99120; GB:AL009126; NID:G2635613; PIDN:CAB15183.1; PID:G2635690

Experimental source: strain 168

Genetics:

Gene: yuku

Superfamily: Bacillus subtilis hypothetical protein yukuJ

Query Match 63.9%; Score 39; DB 2; Length 145;

Best Local Similarity 66.7%; Pred. No. 31;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWVV 9

|||

DB 6 KWHWII 11

RESULT 13

C75311

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

Species: Deinococcus radiodurans

Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

Accession: C75311

R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

Reference number: A75250; MUID:20036896; PMID:10567266

Accession: C75311

Status: preliminary

Molecule type: DNA

Residues: 1-159 <WHI>

Cross-references: GB:AE002048; GB:AE000513; NID:G6459929; PIDN:RAF11686.1; PID:G645993

Experimental source: strain R1

Genetics:

Gene: DR2142

Map position: 1

Query Match 63.9%; Score 39; DB 2; Length 159;

Best Local Similarity 62.5%; Pred. No. 34;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LHKWHVV 9
-| |||
DB 11 VHPWHVV 18

RESULT 14

T06656

hypothetical protein T6G15.90 - Arabidopsis thaliana

Species: Arabidopsis thaliana (mouse-ear cress)

Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999

Accession: T06656

R.; Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 1999
A; Reference number: Z15791
A; Accession: T06656
A; Molecule type: DNA
A; Residues: 1-210 <BEV>

Cross-references: EMBL:AL049656; GSPDB:GN00062; ATSP:T6G15.90

Experimental source: cultivar Columbia; BAC clone T6G15

Genetics:

Gene: ATSP:T6G15.90

Map position: 4

Introns: 122/1

Query Match 63.9%; Score 39; DB 2; Length 210;

Best Local Similarity 71.4%; Pred. No. 45;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWVV 9

|||

DB 79 HEWEWVV 85

RESULT 15

AG2462

hypothetical protein all5255 [imported] - Nostoc sp. (strain PCC 7120)

Species: Nostoc sp. PCC 7120

Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

Accession: AG2462

R.; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001

Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat

Reference number: AB1807; MUID:21595285; PMID:11759840

Accession: AG2462

Status: preliminary

Molecule type: DNA

Residues: 1-259 <KUR>

Cross-references: GB:BA000019; PIDN:BA076954.1; PID:gl7134394; GSPDB:GN00179

Experimental source: strain PCC 7120

Genetics:

Gene: all5255

Query Match 63.9%; Score 39; DB 2; Length 259;

Best Local Similarity 66.7%; Pred. No. 55;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHVV 7

|||

DB 1 MHTWHV 6

Search completed: December 16, 2003, 14:19:16

Job time: 12.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:35 ; Search time 5.66667 Seconds
(without alignments)
74.689 Million cell updates/sec

Title: US-09-870-089B-7
Perfect score: 61
Sequence: 1 FLKWHWV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	68.9	172	1 VIF_SIVGB	P22383 simian immu
2	40	65.6	88	1 TATA_MYCLE	P54079 mycobacteri
3	39	63.9	143	1 RL15_MRIJA	P54047 methanococ
4	39	63.9	310	1 YL60_BUCAP	Q8k9x4 buchnera ap
5	39	63.9	360	1 IL8B_BOVIN	Q28003 bos taurus
6	39	63.9	529	1 C6T1_DROME	Q9vri9 drosophila
7	38	62.3	79	1 TATA_HELPJ	Q9zmb8 helicobacte
8	38	62.3	79	1 TATA_HELPY	Q25088 helicobacte
9	38	62.3	158	1 YBHE_ECOLI	P12994 escherichia
10	38	62.3	179	1 YH94_STRCO	Q9x9z8 streptomyce
11	38	62.3	367	1 YC36_HABIN	P44132 haemophilus
12	38	62.3	380	1 COSC_YEAST	P53053 saccharomyc
13	38	62.3	628	1 HCVC_EURCA	Q9nfl6 eurytelma c
14	38	62.3	630	1 HCVA_EURCA	P14750 eurytelma c
15	37	60.7	138	1 YB27_HABIN	Q86234 haemophilus
16	37	60.7	153	1 MRGA_BACSU	P37960 bacillus su
17	37	60.7	339	1 SP5G_BACSU	P39627 bacillus su
18	37	60.7	389	1 ALKB_HUMAN	Q13686 homo sapien
19	37	60.7	623	1 HCVE_EURCA	P02242 eurytelma c
20	37	60.7	626	1 HCY6_ANDAU	P80476 androctonus
21	37	60.7	626	1 HCVE_EURCA	Q9nfh9 eurytelma c
22	37	60.7	626	1 HCYD_EURCA	P02241 eurytelma c
23	37	60.7	628	1 HCY2_LIMPO	P04253 limulus pol
24	37	60.7	628	1 HCYF_EURCA	Q9nfl5 eurytelma c
25	37	60.7	965	1 PSD1_CAEEL	Q18115 caenorhabdi
26	37	60.7	1002	1 VGNM_CPSMV	P31630 cowpea seve
27	36.5	59.8	220	1 NADD_YERPE	Q82891 yersinia pe
28	36.5	59.8	5596	1 MDNI_HUMAN	Q9nu22 homo sapien
29	36	59.0	144	1 G20U_BACSU	P80879 bacillus su
30	36	59.0	240	1 CYBH_AZOCH	Q43953 azotobacter
31	36	59.0	240	1 CYBH_AZOVI	P23000 azotobacter
32	36	59.0	299	1 BAH_STRHY	Q01109 streptomyce
33	36	59.0	380	1 YCX7_ODOSI	P49833 odontella s

ALIGNMENTS

RESULT 1

VIF_SIVGB STANDARD; PRT; 172 AA.
AC P22383;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Virion infectivity factor (SOR protein) (Q protein).
GN VIF.
OS Simian immunodeficiency virus (isolate GBI).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11732;
RN [1]_TaxID=11732;
RP SEQUENCE FROM N.A.
RX MEDLINE=90015168; PubMed=2797181;
RA Tsujimoto H., Hasegawa A., Maki N., Fukasawa M., Miura T., Speidel S.,
RA Cooper R.W., Moriama E.N., Gojobori T., Hayami M.;
RT "Sequence of a novel simian immunodeficiency virus from a wild-caught
RT African mandrill";
RL Nature 341:539-541(1989).
CC -!- FUNCTION: DETERMINES VIRUS INFECTIVITY.
CC -!- MISCELLANEOUS: THIS IS AN AFRICAN MANDRILL ISOLATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).

DR EMBL; M27470; AAB49570.1; -
DR InterPro; IPR000475; Viral_infect.
DR Pfam; PF00559; Vir_1.
DR PRINTS; PR00349; VIRIONINFCT.
DR ProDom; PD000064; Viral_infect; 1.
KW AIDS.

SQ SEQUENCE 172 AA; 20499 MW; 205E7BDE6265FCB CRC64;

Query Match 68.9%; Score 42; DB 1; Length 172;

Best Local Similarity 62.5%; Pred. No. 8.6;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWV 9
: |||||
Db 19 IERKHWLV 26

RESULT 2

TATA_MYCLE STANDARD; PRT; 88 AA.
ID TATA_MYCLE
AC P54079;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sec-independent protein translocase protein tata/E homolog.

```
GN TATA OR ML1331 OR MLCB2533.27 OR U2126B OR B2126_C1_182.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinometales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: REQUIRED FOR CORRECT LOCALIZATION OF PRECURSOR PROTEINS
CC BEARING SIGNAL PEPTIDES WITH THE TWIN ARGININE CONSERVED MOTIF
CC S/T-R-X-F-L-K. THIS SEC-INDEPENDENT PATHWAY IS TERMED TAT FOR
CC TWIN-ARGININE TRANSLLOCATION SYSTEM. THIS SYSTEM MAINLY TRANSPORTS
CC PROTEINS WITH BOUND COFACTORS THAT REQUIRE FOLDING PRIOR TO EXPORT
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -!- SIMILARITY: BELONGS TO THE TATA/E FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00017; AA117190.1; -.
CC EMBL; AL035310; CAA22941.1; -.
CC EMBL; AL583921; CAC31712.1; ALT_INIT.
CC PIR; S72850; S72850.
CC Leproma; ML1331; -.
CC HAMAP; MF_00236; -.
CC InterPro; IPR003369; Meta_Hcf106.
CC InterPro; IPR006312; TATA_E.
CC Pfam; PF02416; Meta_Hcf106; 1.
CC TIGRFAMs; TIGR01411; tataE; 1.
CC Transport; Protein transport; Translocation; Transmembrane;
CC Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
SQ SEQUENCE 88 AA; 9824 MW; 97F8C81009F7CAC8 CRC64;
Query Match 65.6%; Score 40; DB 1; Length 88;
Best Local Similarity 75.0%; Pred. No. 9.2;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 LHKWHWV 9
DB 4 LSPWHWV 11
RESULT 3
RL15_METJA
ID RL15_METJA STANDARD; PRT; 143 AA.
AC P54047;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L15P.
GN RPL15P OR MJ0477.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=21190;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=9688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Kirkness E.F., Tomb J.-F., Weisskopf K.G., Merrick J.M., Glodek A.,
RA Overbeek R., Kirkness N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Scott J.L., Geoghagan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67497; AAB98466.1; -.
CC PIR; E64359; E64359.
CC TIGR; MJ0477; -.
CC InterPro; IPR001196; Ribosomal_L15.
CC Pfam; PF00256; L15; 1.
CC PROSITE; PS00475; RIBOSOMAL_L15; 1.
CC Ribosomal protein; Complete proteome.
SQ SEQUENCE 143 AA; 16107 MW; 2A78A8ABA578FF87 CRC64;
Query Match 63.9%; Score 39; DB 1; Length 143;
Best Local Similarity 57.1%; Pred. No. 20;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 HKWHWV 9
DB 42 HKWTWII 48
RESULT 4
Y160_BUCAP
ID Y160_BUCAP STANDARD; PRT; 310 AA.
AC Q8K9X4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein Buegl160.
GN BUEGL160.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamas L., Klaesson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria."
RL Science 296:2376-2379(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE TERC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; A5014091; AM67728.1; -;
 DR InterPro; IPR005496; Terc.
 DR Pfam; PF03741; Terc; 1.

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 74 94 POTENTIAL.
 FT TRANSMEM 110 130 POTENTIAL.
 FT TRANSMEM 135 155 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 228 248 POTENTIAL.
 FT TRANSMEM 256 276 POTENTIAL.
 FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 310 AA; 36938 MW; 44B75CDED6B47004 CRC64;
 SQ SEQUENCE

Query Match 63.9%; Score 39; DB 1; Length 310;

Best Local Similarity 66.7%; Pred. No. 41;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHVV 9

|||||

Db 132 KHWIL 137

RESULT 5

IL8B_BOVIN

ID IL8B_BOVIN STANDARD; PRT; 360 AA.

AC Q28003;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE High affinity interleukin-8 receptor B (IL-8R B) (CXCR-2).

GN IL8RB OR CXCR2.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OC NCBI_TaxID=9913;

OX [1]

RP SEQUENCE FROM N.A.

RA Li Y., Feng J., Templeton J.W.;

RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: RECEPTOR TO INTERLEUKIN-8, WHICH IS A POWERFUL

CC NEUTROPHILS CHEMOTACTIC FACTOR. BINDING OF IL-8 TO THE RECEPTOR

CC CAUSES ACTIVATION OF NEUTROPHILS. THIS RESPONSE IS MEDIATED VIA A

CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND

CC MESSENGER SYSTEM. THIS RECEPTOR BINDS TO IL-8 WITH A HIGH AFFINITY

CC AND TO GRO/MGSA AND NAP-2 ALSO WITH A HIGH AFFINITY (BY

CC SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U19947; AAA8496.1; -;

DR InterPro; IPR000276; GPCR_Rhodpsn.

DR Pfam; PF00001; 7tm.1; 1.

DR PRINTS; PR00237; GPCRHHODPSN.

DR PROSITE; PS00237; G PROTEIN RECEPTOR F1.1; 1.

DR PROSITE; PS00262; G-PROTEIN RECEPTOR F1.2; 1.

CC G-protein coupled receptor; Transmembrane; Glycoprotein;

KW Chemotaxis.
 FT DOMAIN 1 48 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 49 75 1 (POTENTIAL).
 FT DOMAIN 76 84 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 85 105 2 (POTENTIAL).
 FT DOMAIN 106 120 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 121 142 3 (POTENTIAL).
 FT DOMAIN 143 163 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 164 183 4 (POTENTIAL).
 FT DOMAIN 184 208 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 209 231 5 (POTENTIAL).
 FT DOMAIN 232 251 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 252 273 6 (POTENTIAL).
 FT DOMAIN 274 294 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 295 315 7 (POTENTIAL).
 FT DOMAIN 316 360 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 119 196 BY SIMILARITY.
 FT CARBOHYD 10 10 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 360 AA; 40625 MW; 9A7F0C982A632D1 CRC64;

Query Match 63.9%; Score 39; DB 1; Length 360;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KHWV 8

|||||

Db 158 KHWV 162

RESULT 6

C6T1_DROME

ID C6T1_DROME STANDARD; PRT; 529 AA.

AC Q9VRI9;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable cytochrome P450 6t1 (EC 1.14.-.-) (CYP6T1).

GN CYP6T1 OR CG1644.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OC NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Betsch M.R., Bertan B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dou L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Posler A., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harria N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalaali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RA "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC (Potential).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF003568; AAF50807.1; -.
CC HSSP; P14779; IJPZ.
CC FlyBase; FBgn031182; Cyp6t1.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00866; CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum; Hypothetical protein.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 529 AA; 60518 MW; 2BAD7875F3BD9C10 CRC64;

Query Match 63.9%; Score 39; DB 1; Length 529;
Best Local Similarity 83.3%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHW 7
DB 41 LHFWEW 46

RESULT 7
TATA_HELPY
ID TATA_HELPY STANDARD; PRT; 79 AA.
AC Q9ZMB8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sec-independent protein translocase protein tatA/E homolog.
GN TATA OR JHP0303.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.B., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen *Helicobacter pylori*.";

RL Nature 397:176-180(1999).
CC -!- FUNCTION: REQUIRED FOR CORRECT LOCALIZATION OF PRECURSOR PROTEINS
CC BEARING SIGNAL PEPTIDES WITH THE TWIN ARGININE CONSERVED MOTIF
CC S/T-R-X-F-L-K. THIS SEC-INDEPENDENT PATHWAY IS TERMED TAT FOR
CC TWIN-ARGININE TRANSLLOCATION SYSTEM. THIS SYSTEM MAINLY TRANSPORTS
CC PROTEINS WITH BOUND COFACTORS THAT REQUIRE FOLDING PRIOR TO EXPORT
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Inner-membrane bound (Probable).
CC -!- SIMILARITY: BELONGS TO THE TATA/E FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE001467; AAD05888.1; -.
CC PIR; A71948; A71948.
CC HAMAP; MF 00236; -; 1.
CC InterPro; IPR003369; MtaA_Hcf106.
CC InterPro; IPR006312; TATA_E.
CC Pfam; PF02416; MtaA_Hcf106; 1.
CC TIGRFAMs; TIGR01411; tataE; 1.
CC Transport; Protein transport; Translocation; Transmembrane;
KW Inner membrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
SQ SEQUENCE 79 AA; 8752 MW; D39E9DCA02AA5E8F CRC64;

Query Match 62.3%; Score 38; DB 1; Length 79;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FLHKWHWV 9
DB 4 FTSIWHWVI 12

RESULT 8
TATA_HELPY
ID TATA_HELPY STANDARD; PRT; 79 AA.
AC Q25088;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sec-independent protein translocase protein tatA/E homolog.
GN TATA OR HP0320.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen *Helicobacter*
RT *pylori*.";
RL Nature 388:539-547(1997).
CC -!- FUNCTION: REQUIRED FOR CORRECT LOCALIZATION OF PRECURSOR PROTEINS
CC BEARING SIGNAL PEPTIDES WITH THE TWIN ARGININE CONSERVED MOTIF
CC S/T-R-X-F-L-K. THIS SEC-INDEPENDENT PATHWAY IS TERMED TAT FOR
CC TWIN-ARGININE TRANSLLOCATION SYSTEM. THIS SYSTEM MAINLY TRANSPORTS
CC PROTEINS WITH BOUND COFACTORS THAT REQUIRE FOLDING PRIOR TO EXPORT

```

CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Inner-membrane bound (Probable).
CC -1- SIMILARITY: BELONGS TO THE TATA/E FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000550; AAD07397.1; -.
DR PIR; H64559; H64559.
DR TIGR; HP0320; -.
DR HAMAP; MF_00236; -.
DR InterPro; IPR003369; MtaA_Hcf106.
DR InterPro; IPR006312; TATA_E.
DR Pfam; PF02416; MtaA_Hcf106; 1.
DR TIGRFAMs; TIGR01411; tataE; 1.
KW Transport; Protein transport; Translocation; Transmembrane;
KW Inner membrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
SQ SEQUENCE 79 AA; 8722 MW; CE599ACA02AA5E8F CRC64;

Query Match 62.3%; Score 38; DB 1; Length 79;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FLHKHWV 9
Db | |||||
4 FTSIHWVI 12

RESULT 9
YBHB ECOLI
ID YBHB_ECOLI STANDARD; PRT; 158 AA.
AC F12994; P75766;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein ybhb.
GN YBHB OR B0773.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89066784; PubMed=3058702;
RA Otsuka A.J., Buoncristiani M.R., Howard P.K., Flamm J., Johnson O.,
RA Yamamoto R., Uchida K., Cook C., Ruppert J., Matsuzaki J.;
RT "The Escherichia coli biotin biosynthetic enzyme sequences predicted
RT from the nucleotide sequence of the bio operon.";
RL J. Biol. Chem. 263:19577-19585(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KL12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:12453-12474(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.66 ANGSTROMS), AND CHARACTERIZATION.
RX MEDLINE=21332775; PubMed=11439028;
RA Serre L., Pereira de Jesus K., Zelwer C., Bureaud N., Schoentgen F.,
RA Benedetti H.;
RT "Crystal structures of YBHB and YBCL from Escherichia coli, two
RT bacterial homologues to a Raf kinase inhibitor protein.";
RL J. Mol. Biol. 310:617-634(2001).

```

```

CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE UPF0098 FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04423; AAA33513.1; -.
DR EMBL; AE000180; AAC73860.1; -.
DR PIR; E64813; Q3ECEA.
DR PDB; 1FJQ; 18-JUL-01.
DR EcoGene; EG11238; ybhb.
DR InterPro; IPR005247; Cons_hypoth481.
DR InterPro; IPR001858; PBP.
DR Pfam; PF01161; PBP; 1.
DR TIGRFAMs; TIGR00481; TIGR00481; 1.
KW 3D-structure; Complete proteome.
FT CONFLICT 136 136 G -> A (IN REF. 1).
FT STRAND 2 4
FT TURN 10 11
FT STRAND 13 13
FT HELIX 16 18
FT STRAND 19 19
FT TURN 23 24
FT STRAND 34 37
FT TURN 41 42
FT STRAND 45 52
FT TURN 53 54
FT TURN 56 59
FT STRAND 61 70
FT TURN 71 72
FT STRAND 75 76
FT TURN 78 79
FT HELIX 80 82
FT TURN 83 83
FT TURN 89 90
FT STRAND 92 93
FT TURN 95 95
FT TURN 97 98
FT STRAND 101 101
FT STRAND 106 106
FT TURN 110 111
FT STRAND 114 123
FT TURN 132 133
FT HELIX 136 146
FT STRAND 147 158
SQ SEQUENCE 158 AA; 17085 MW; 0AD6F21157060A4C CRC64;

Query Match 62.3%; Score 38; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 WHWV 9
Db | |||||
62 WHWV 66

RESULT 10
YH94 STRCO
ID YH94_STRCO STANDARD; PRT; 179 AA.
AC Q9X9Z8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein SCO1794.
GN SCO1794 OR SC15.02.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

```

```

OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Lake L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- SIMILARITY: BELONGS TO THE UPF0098 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL939110; CAB45286.1; -.
DR PIR; T36855; T36855.
DR HSSP; P77368; 1FUX.
DR InterPro; IPR005247; Cons_hypoth481.
DR InterPro; IPR001858; PBP.
DR Pfam; PF01161; PBP; 1.
DR TIGRFAMs; TIGR00481; TIGR00481; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 179 AA; 19158 MW; ACA6A43BD830861 CRC64;

Query Match 62.3%; Score 38; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 WHWV 9
Db 76 WHWV 80

RESULT 11
YC36 HAEIN
ID YC36 HAEIN STANDARD; PRT; 367 AA.
AC P44132;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein H11236.
GN H11236.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;

"Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: TO E.COLI YDGA AND YIHf.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32803; AAC22888.1; -.
DR PIR; F64022; F64022.
DR TIGR; H11236; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
SQ SEQUENCE 367 AA; 41355 MW; B00211C5F98C2585 CRC64;

Query Match 62.3%; Score 38; DB 1; Length 367;
Best Local Similarity 80.0%; Pred. No. 67;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KWHW 8
Db 363 KWHW 367

RESULT 12
COSC YEAST
ID COSC YEAST STANDARD; PRT; 380 AA.
AC P53053;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE COS12 protein.
GN COS12 OR YGL263W OR NRC380.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97127827; PubMed=8972578;
RA Coissac E., Maillier E., Robineau S., Netter P.;
RT "Sequence of a 39,411 bp DNA fragment covering the left end of
RL chromosome VII of Saccharomyces cerevisiae.";
RL Yeast 12:1553-1562(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE DUP/COS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X94357; CAA64124.1; -.
DR EMBL; Z72785; CAA96983.1; -.
DR PIR; S61598; S61598.
DR SGD; S0003232; COS12.
DR InterPro; IPR001142; DUP.
DR Pfam; PF00674; DUP; 2.
KW Transmembrane; Multigene family.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 232 252 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
SQ SEQUENCE 380 AA; 44803 MW; 52F58F4E400CA596 CRC64;

```

```
Query Match 62.3%; Score 38; DB 1; Length 380;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 KHWV 8
Db 194 KHWI 198

RESULT 13
HCYC_EURCA STANDARD; PRT; 628 AA.
AC Q9NFI6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemocyanin C chain (HcC)
GN HCC.
OS Eurytelma californica (American tarantula).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Aphonopelma.
OX NCBI_TaxID=29932;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=20564303; PubMed=10961996;
RA Voit R., Feldmaier-Fuchs G., Schweikard T., Decker H., Burmester T.;
RT "Complete sequence of the 24-mer hemocyanin of the tarantula Eurytelma
californicum. Structure and intramolecular evolution of the
subunits."
RL J. Biol. Chem. 275:39339-39344 (2000).
CC -!- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
ARTHROPODS.
CC -!- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN
DIFFERENT CHAINS IDENTIFIED.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Hemolymph.
CC -!- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN
LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A
BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN
ADDITION TO BINDING OXYGEN.
CC -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ277489; CAB89495.1; -.
CC HSSP; P04253; 1LLA.
CC InterPro; IPR000896; Hemocyanin.
CC InterPro; IPR005203; hemocyanin_C.
CC InterPro; IPR005204; hemocyanin_N.
CC InterPro; IPR002227; Tyrosinase.
CC Pfam; PF03722; hemocyanin; 1.
CC Pfam; PF03723; hemocyanin_C; 1.
CC PROSITE; PS00209; HEMOCYANIN; 1.
CC PROSITE; PS00210; HEMOCYANIN; 1.
CC PROSITE; PS00498; TYROSINASE; 2; 1.
CC PROSITE; PS00499; TYROSINASE; 2; 1.
CC Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.
KW INIT MET 0 BY SIMILARITY.
FT DISULFID 536 584 BY SIMILARITY.
FT METAL 174 174 COPPER 1 (BY SIMILARITY).
FT METAL 178 178 COPPER 1 (BY SIMILARITY).
FT METAL 205 205 COPPER 1 (BY SIMILARITY).
FT METAL 325 325 COPPER 2 (BY SIMILARITY).
FT METAL 329 329 COPPER 2 (BY SIMILARITY).
FT METAL 329 329 COPPER 2 (BY SIMILARITY).

Query Match 62.3%; Score 38; DB 1; Length 628;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 HKWHV 9
Db 173 HKWHV 179

RESULT 14
HCYA_EURCA STANDARD; PRT; 630 AA.
AC P14750;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemocyanin A chain (HcA)
GN HCA.
OS Eurytelma californica (American tarantula).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Aphonopelma.
OX NCBI_TaxID=29932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91060544; PubMed=2246235;
RA Voit R., Feldmaier-Fuchs G.;
RT "Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs
encoding the tarantula hemocyanin subunits a and e."
RL J. Biol. Chem. 265:19447-19452 (1990).
RN [2]
RP SEQUENCE
RX MEDLINE=91025623; PubMed=2222854;
RA Schartau W., Metzger W., Sonner P., Geisert H., Storz H.;
RT "Hemocyanins in spiders, XXII. Complete amino-acid sequence of
subunit a of Eurytelma californicum hemocyanin."
RL Biol. Chem. Hoppe-Sevler 371:557-565 (1990).
CC -!- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
ARTHROPODS.
CC -!- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN
DIFFERENT CHAINS IDENTIFIED.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Hemolymph.
CC -!- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN
LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A
BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN
ADDITION TO BINDING OXYGEN.
CC -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X16893; CAA34771.1; -.
CC HSSP; P04253; 1LLA.
CC InterPro; IPR000896; Hemocyanin.
CC InterPro; IPR005203; hemocyanin_C.
CC InterPro; IPR005204; hemocyanin_N.
CC InterPro; IPR002227; Tyrosinase.
CC Pfam; PF03722; hemocyanin; 1.
CC Pfam; PF03723; hemocyanin_C; 1.
CC PROSITE; PS00209; HEMOCYANIN; 1.
CC PROSITE; PS00210; HEMOCYANIN; 1.
CC PROSITE; PS00498; TYROSINASE; 2; 1.
CC PROSITE; PS00499; TYROSINASE; 2; 1.
CC Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.
KW INIT MET 0 BY SIMILARITY.
FT DISULFID 536 584 BY SIMILARITY.
FT METAL 174 174 COPPER 1 (BY SIMILARITY).
FT METAL 178 178 COPPER 1 (BY SIMILARITY).
FT METAL 205 205 COPPER 1 (BY SIMILARITY).
FT METAL 325 325 COPPER 2 (BY SIMILARITY).
FT METAL 329 329 COPPER 2 (BY SIMILARITY).
```

```
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
KW Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.
FT INIT MET 0
FT METAL 174 174 COPPER 1 (PROBABLE).
FT METAL 178 178 COPPER 1 (PROBABLE).
FT METAL 205 205 COPPER 1 (PROBABLE).
FT METAL 325 325 COPPER 2 (PROBABLE).
FT METAL 329 329 COPPER 2 (PROBABLE).
FT METAL 365 365 COPPER 2 (PROBABLE).
FT METAL 365 365 COPPER 2 (PROBABLE).
FT DISULFID 537 585 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 450 450 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 630 AA; 72187 MW; 4F69899631CC793 CRC64;

Query Match 62.3%; Score 38; DB 1; Length 630;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 HKWHWV 9
DB 173 HHWHHV 179

RESULT 15
YB27 HAEIN STANDARD; PRT; 138 AA.
AC O86234,
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein HI1127.
GN HI1127.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; Pubmed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D., Peterson J.,
RA Hickey E., Dodson R., Gwinn M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32792; AAC22782.1; -.
DR TIGR; HI1127; -.

KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
SQ SEQUENCE 138 AA; 15598 MW; 8833E1195A6184EC CRC64;

Query Match 60.7%; Score 37; DB 1; Length 138;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HKWHWV 8
DB 109 HKFHWV 114

Search completed: December 16, 2003, 14:15:15
Job time : 6.6667 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:09:45 ; Search time 25 Seconds
(without alignments)
92.899 Million cell updates/sec

Title: US-09-870-089B-7
Perfect score: 61
Sequence: 1 FLHKWHVVV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	46	75.4	209	16 Q9A3Z1	Q9A3Z1 caulobacter
2	46	75.4	213	16 Q8P1F4	Q8P1F4 xanthomonas
3	46	75.4	214	16 Q8P739	Q8P739 xanthomonas
4	46	75.4	218	16 Q9PFR0	Q9PFR0 xylella fas
5	46	75.4	328	10 Q9ZR87	Q9ZR87 zinnia eleg
6	46	75.4	417	2 Q8VM89	Q8VM89 rhizobium l
7	43	70.5	196	16 Q8RHF3	Q8RHF3 fusobacteri
8	42	68.9	486	16 Q8E8L4	Q8E8L4 shewanella
9	42	68.9	737	17 Q8ZXG1	Q8ZXG1 pyrobaculum
10	41	67.2	145	5 Q8MPP4	Q8MPP4 caenorhabdi
11	41	67.2	294	8 Q9XMG3	Q9XMG3 epichloe ty
12	41	67.2	573	2 Q8KQT4	Q8KQT4 xanthomonas
13	41	67.2	573	16 Q8P338	Q8P338 xanthomonas
14	41	67.2	574	16 Q8PEH7	Q8PEH7 xanthomonas
15	41	67.2	582	16 Q92WU1	Q92WU1 rhizobium m
16	41	67.2	670	5 Q9BLD9	Q9BLD9 drosophila

17	41	67.2	681	5	077002	077002 hyphantria
18	41	67.2	683	5	Q95R43	Q95R43 drosophila
19	41	67.2	683	5	Q9W1V6	Q9W1V6 drosophila
20	41	67.2	683	5	Q964D5	Q964D5 galliera me
21	41	67.2	683	5	Q76951	Q76951 anopheles g
22	41	67.2	683	5	Q811F6	Q811F6 drosophila
23	41	67.2	684	5	Q96752	Q96752 anopheles g
24	41	67.2	684	5	Q9W521	Q9W521 drosophila
25	41	67.2	684	5	Q9GYW1	Q9GYW1 aedes aegyp
26	41	67.2	684	5	Q97047	Q97047 tenebrio mo
27	41	67.2	684	5	Q96751	Q96751 anopheles g
28	41	67.2	684	5	Q9N2P8	Q9N2P8 armigeres s
29	41	67.2	684	5	Q816K2	Q816K2 holotrichia
30	41	67.2	684	5	Q816K1	Q816K1 holotrichia
31	41	67.2	685	5	Q44249	Q44249 manduca sex
32	41	67.2	685	5	Q8MZM2	Q8MZM2 anopheles g
33	41	67.2	685	5	Q9GYW2	Q9GYW2 aedes aegyp
34	41	67.2	685	5	Q9GU89	Q9GU89 bombyx mori
35	41	67.2	685	5	Q910B4	Q910B4 sarcophaga
36	41	67.2	685	5	Q27451	Q27451 bombyx mori
37	41	67.2	686	5	Q8WSA8	Q8WSA8 aedes aegyp
38	41	67.2	686	5	Q44251	Q44251 anopheles g
39	41	67.2	686	5	Q96453	Q96453 anopheles s
40	41	67.2	687	5	Q96753	Q96753 anopheles g
41	41	67.2	690	5	Q9GVA6	Q9GVA6 pimpla hypo
42	41	67.2	695	5	Q25519	Q25519 manduca sex
43	41	67.2	696	5	Q8MZM4	Q8MZM4 anopheles g
44	41	67.2	697	5	Q76208	Q76208 hyphantria
45	41	67.2	699	5	Q9GVA5	Q9GVA5 pimpla hypo

ALIGNMENTS

RESULT 1

Q9A3Z1 PRELIMINARY; PRT; 209 AA.
ID Q9A3Z1
AC Q9A3Z1;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)
DE Hypothetical protein CC3059.
GN CC3059.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Bisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Salzberg S.L., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA "Complete genome sequence of Caulobacter crescentus."
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AB005969; AAK25021.1; -.
DR TIGR; CC3059; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 209 AA; 22817 MW; B8C635B8A3643259 CRC64;

Query Match 75.4%; Score 46; DB 16; Length 209;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 LHKWHVV 8
DB 17 LHKWHVV 23

```

RESULT 2
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.F., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
"Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AE012390; AAM42047.1; -.
DR InterPro; IPR005797; Cytb_b6_N.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 214 AA; 23740 MW; 0F7BFDCC76937150 CRC64;

Query Match 75.4%; Score 46; DB 16; Length 214;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LHKWHWV 8
Db 24 LHQWHWI 30
|||||
|||||

RESULT 4
Q9PFR0 PRELIMINARY; PRT; 218 AA.
ID Q9PFR0 AC Q9PFR0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein Xf0597.
GN XF0597.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.B., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
"The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).

```

```
DR EMBL; AE003905; AAF83407.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 218 AA; 24198 MW; A032E2AD602C24B8 CRC64;

Query Match 75.4%; Score 46; DB 16; Length 218;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKWHWV 8
DB 28 LHKWHWV 34

RESULT 5
Q9ZR87 PRELIMINARY; PRT; 328 AA.
AC Q9ZR87
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bifunctional nuclease.
GN NUC2E2.
OS Zinnia elegans.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Zinnia.
OX NCBI_TaxID=34245;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20098724; PubMed=10631260;
RA Perez-Amador M.A., Adler M.D., De Rocher E.J., Thompson D.M.,
RA van Hoof A., LeBrasseur N.D., Lers A., Green P.J.;
RT "Identification of BFN1, a bifunctional nuclease induced during leaf
RT and stem senescence in arabidopsis.";
RL Plant Physiol. 122:169-179(2000).
DR EMBL; U90266; AAD00695.1; -.
DR HSP; F24289; IAKO.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR003154; SI/Plnuclease.
DR Pfam; PF02265; Nuclease; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
SQ SEQUENCE 328 AA; 37560 MW; 9343388936C53F7E CRC64;

Query Match 75.4%; Score 46; DB 10; Length 328;
Best Local Similarity 71.4%; Pred. No. 18;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKWHW 7
DB 75 WMHKWHW 81

RESULT 6
Q8VM89 PRELIMINARY; PRT; 417 AA.
AC Q8VM89
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Two-component sensor.
GN MCTS.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3841;
RA Hosie A.H.F., Poole P.S.;
RT "Identification and characterisation of a novel monocarboxylic acid
RT transporter of Rhizobium leguminosarum.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ421944; CAD19126.1; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR Pfam; PF02518; HATPase_c; 1.
DR SMART; SM00387; HATPase_c; 1.
SQ SEQUENCE 417 AA; 45820 MW; 2034030279B9E996 CRC64;

Query Match 75.4%; Score 46; DB 2; Length 417;
Best Local Similarity 87.5%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWV 9
DB 124 LDKWHWV 131

RESULT 7
Q8RHF3 PRELIMINARY; PRT; 196 AA.
AC Q8RHF3
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein FN2083.
GN FN2083.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasleva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AB010511; AAL94167.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 196 AA; 23169 MW; 4438206B08CCB0C7 CRC64;

Query Match 70.5%; Score 43; DB 16; Length 196;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWV 9
DB 146 LDKWHWV 153

RESULT 8
Q8E8L4 PRELIMINARY; PRT; 486 AA.
AC Q8E8L4
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SC4649.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
```



```
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vanatavean J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015896; AAN57609.1; -.
DR TIGR; SO4649; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 486 AA; 57471 MW; B25617A4A982564F CRC64;

Query Match 68.9%; Score 42; DB 16; Length 486;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKWHWV 9
Db 4 LHRWNWLV 11
|:|:|:|
|:|:|:|

RESULT 9
Q8ZXG1 PRELIMINARY; PRT; 737 AA.
AC Q8ZXG1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Dipeptide binding protein.
GN PAI300.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum."
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009814; AAL63387.1; -.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 3.
KW Complete proteome.
SQ SEQUENCE 737 AA; 82378 MW; BCF9D3E46354725 CRC64;

Query Match 68.9%; Score 42; DB 17; Length 737;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLHKWHWV 8
Db 423 FHRQNSWV 430
|:|:|:|
|:|:|:|

RESULT 10
Q8MPP4 PRELIMINARY; PRT; 145 AA.
AC Q8MPP4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Y46G5A.33 protein.
GN Y46G5A.33
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]

RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vanatavean J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015896; AAN57609.1; -.
DR TIGR; SO4649; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 486 AA; 57471 MW; B25617A4A982564F CRC64;

Query Match 68.9%; Score 42; DB 16; Length 486;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKWHWV 9
Db 4 LHRWNWLV 11
|:|:|:|
|:|:|:|

RESULT 9
Q8ZXG1 PRELIMINARY; PRT; 737 AA.
AC Q8ZXG1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Dipeptide binding protein.
GN PAI300.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum."
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009814; AAL63387.1; -.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 3.
KW Complete proteome.
SQ SEQUENCE 737 AA; 82378 MW; BCF9D3E46354725 CRC64;

Query Match 68.9%; Score 42; DB 17; Length 737;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLHKWHWV 8
Db 423 FHRQNSWV 430
|:|:|:|
|:|:|:|

RESULT 10
Q8MPP4 PRELIMINARY; PRT; 145 AA.
AC Q8MPP4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Y46G5A.33 protein.
GN Y46G5A.33
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]

RP SEQUENCE FROM N.A.
RA Wallis J.M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL110485; CAD36500.1; -.
SQ SEQUENCE 145 AA; 16489 MW; 50B4B1F300F7D357 CRC64;

Query Match 67.2%; Score 41; DB 5; Length 145;
Best Local Similarity 50.0%; Pred. No. 47;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLHKWHWV 8
Db 94 FHRFHWI 101
|:|:|:|
|:|:|:|

RESULT 11
Q9XMG3 PRELIMINARY; PRT; 294 AA.
AC Q9XMG3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Et2.OL ORF1 protein (Fragment).
GN Et2.OL ORF1.
OS Epichloe typhina.
OG Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Clavicipitaceae; Epichloe.
OX NCBI_TaxID=5113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PRG;
RX MEDLINE=92145777; PubMed=1782678;
RA Mogen K.L., Siegel M.R., Schardl C.L.;
RT "Linear DNA plasmids of the perennial ryegrass choke pathogen Epichloe
RT typhina (Clavicipitaceae).";
RL Curr. Genet. 20:519-526(1991).
DR EMBL; X57200; CAA40486.2; -.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase; Transferase; Mitochondrion; Plasmid.
FT NON_TER 1 294
FT NON_TER 1 294
SQ SEQUENCE 294 AA; 35780 MW; B59525CA6962B05E CRC64;

Query Match 67.2%; Score 41; DB 8; Length 294;
Best Local Similarity 71.4%; Pred. No. 92;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLHKWHWV 7
Db 57 FTHKWNW 63
|:|:|:|
|:|:|:|

RESULT 12
Q8KQT4 PRELIMINARY; PRT; 573 AA.
AC Q8KQT4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 60kd inner-membrane protein.
GN Y1DC.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
```

```

OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22034966; PubMed=12039751;
RA Yen M.R., Lin N.T., Hung C.H., Choy K.T., Weng S.F., Teeng Y.H.;
RT "oriC Region and Replication Termination Site, dif, of the Xanthomonas
campestris pv. campestris 17 Chromosome.";
RL Appl. Environ. Microbiol. 68:2924-2933 (2002).
DR EMBL; AY057934; AAL30086.1; -.
DR InterPro; IPR001708; 60kDa innermem.
DR Pfam; PF02096; 60KD IMP; 1.
DR PRINTS; PR00701; 60KIDINNERMP.
SQ SEQUENCE 573 AA; 63334 MW; 1A2B72162135DFE8 CRC64;

Query Match 67.2%; Score 41; DB 2; Length 573;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FLHKHWVV 9
DB 375 FLHNWGWA 383

RESULT 13
Q8P338 PRELIMINARY; PRT; 573 AA.
AC Q8P338;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 60 kDa inner-membrane protein.
GN XCC4240.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33913 / NCPPB 528;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarotte G., Cannavan F., Cardozo J., Chamberg F., Ciapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities.";
RL Nature 417:459-463 (2002).
DR EMBL; AE012552; AAM43456.1; -.
DR InterPro; IPR001708; 60kDa innermem.
DR Pfam; PF02096; 60KD IMP; 1.
DR PRINTS; PR00701; 60KIDINNERMP.
KW Complete proteome.
SQ SEQUENCE 574 AA; 63297 MW; C3A52D550C85F7AB CRC64;

Query Match 67.2%; Score 41; DB 16; Length 574;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FLHKHWVV 9
DB 376 FLHNWGWA 384

RESULT 15
Q92WU1 PRELIMINARY; PRT; 582 AA.
AC Q92WU1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein RB0238.
GN RB0238 OR SMB20248.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RA MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P., Gouzy J.,
Vorhoelter J.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,

```

RA Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).
DR ENBL; AL603642; CAC48638.1; -. Complete proteome.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 582 AA; 63374 MW; 2E20B8D0DB09B2B CRC64;

Query Match 67.2%; Score 41; DB 16; Length 582;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHVVV 9
||:|:|:
Db 463 LHRWELV 470

Search completed: December 16, 2003, 14:18:00
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:05 ; Search time 33 Seconds
(without alignments)
43.289 Million cell updates/sec

Title: US-09-870-089B-9
Perfect score: 64
Sequence: 1 FLHKWHYV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	64	100.0	9	23 AAU74684	Human cancer antig
2	56	87.5	9	23 AAU74683	Human cancer antig
3	45	70.3	556	22 AAU31978	Novel human secret
4	44	68.8	512	22 ABB57797	Drosophila melanog
5	44	68.8	554	22 AAU14133	Human novel protei
6	44	68.8	683	22 ABB59504	Drosophila melanog
7	44	68.8	684	22 ABB57998	Drosophila melanog
8	44	68.8	685	17 AAR89115	Bombyx mori (pro)p
9	44	68.8	685	18 AAU14441	Prophenol oxidase.

10	43	67.2	142	22 AAU34057	Staphylococcus aur
11	43	67.2	147	22 AAU36776	Staphylococcus aur
12	43	67.2	166	23 ABP39358	Staphylococcus epi
13	43	67.2	235	23 ABP26523	Streptococcus poly
14	43	67.2	1248	22 ABG19706	Novel human diago
15	42	65.6	98	23 ABJ10988	Yeast selected int
16	42	65.6	185	9 AAP82003	Sequence of novel
17	42	65.6	480	16 AAR78268	Bacillus amylioliqu
18	42	65.6	480	17 AAU14499	Bacillus amylioliqu
19	42	65.6	480	20 AAY15419	Bacillus amylioliqu
20	42	65.6	480	20 AAY07385	Wild type Termamyl
21	42	65.6	480	21 AAY95606	Bacillus amylioliqu
22	42	65.6	481	24 ABP60489	Bacillus alpha-amy
23	42	65.6	481	24 ABP60490	Bacillus alpha-amy
24	42	65.6	481	24 ABP60491	Bacillus alpha-amy
25	42	65.6	482	18 AAU08192	B.licheniformis mu
26	42	65.6	482	18 AAU08195	B.licheniformis mu
27	42	65.6	482	18 AAU08196	B.licheniformis mu
28	42	65.6	482	18 AAU08197	B.licheniformis mu
29	42	65.6	482	18 AAU08198	B.licheniformis mu
30	42	65.6	482	18 AAU08199	B.licheniformis mu
31	42	65.6	482	18 AAU08200	B.licheniformis mu
32	42	65.6	482	18 AAU08201	B.licheniformis mu
33	42	65.6	483	12 AAR10576	Mutant alpha-amy
34	42	65.6	483	12 AAR10577	Mutant alpha-amy
35	42	65.6	483	12 AAR10578	Mutant alpha-amy
36	42	65.6	483	12 AAR10579	Mutant alpha-amy
37	42	65.6	483	12 AAR10580	Mutant alpha-amy
38	42	65.6	483	12 AAR10581	Mutant alpha-amy
39	42	65.6	483	12 AAR10582	Mutant alpha-amy
40	42	65.6	483	13 AAR22750	Alpha amylase. Ba
41	42	65.6	483	13 AAR25434	Alpha amylase muta
42	42	65.6	483	13 AAR31362	B.licheniformis al
43	42	65.6	483	14 AAR31363	B.licheniformis al
44	42	65.6	483	14 AAR34288	B.licheniformis al
45	42	65.6	483	14 AAR34289	B.licheniformis al

ALIGNMENTS

RESULT 1
AAU74684

ID AAU74684 standard; Peptide; 9 AA.

XX AC AAU74684;

XX XX

DT 09-APR-2002 (first entry)

DE Human cancer antigen ATF4/CREB-2 based immunogenic ligand #4.

XX Human cancer antigen; ATF4; CREB-2; vaccine; cytostatic;

XX immunogenic ligand; gene therapy; MHC; major histocompatibility complex;

XX adoptive immunotherapy; cancer; ovarian cancer.

OS Homo sapiens.

SS Synthetic.

XX WO200192306-A2.

XX PD 06-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US17454.

XX PR 31-MAY-2000; 2000US-209388P.

XX PR 20-DEC-2000; 2000US-257007P.

XX PA (GENZ) GENZYME CORP.

XX PI Nicolette CA;

XX DR WPI; 2002-097764/13.

XX XX

XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX XX WPI; 2001-611725/70.
DR XX Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
PT PT
XX XX
PS Claim 20; Page 534; 765pp; English.
XX XX
XX The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX XX
XX Sequence 556 AA;
SQ
Query Match 70.3%; Score 45; DB 22; Length 556;
Best Local Similarity 75.0%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 FLHKWHWY 8
DB 74 FLHSMWY 81
RESULT 4
ABBS7797
ID ABB57797 standard; Protein; 512 AA.
XX AC
XX ABB57797;
XX AC
XX 26-MAR-2002 (first entry)
XX DT
DE Drosophila melanogaster polypeptide SEQ ID NO 183.
XX DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX KW
XX Drosophila melanogaster.
XX OS
XX WO200171042-A2.
XX PN
XX 27-SEP-2001.
XX PD
XX 23-MAR-2001; 2001WO-US09231.
XX PF
XX 23-MAR-2000; 2000US-191637P.
XX PR
XX 11-JUL-2000; 2000US-0614150.
XX PR
XX (PEKE) PE CORP NY.
XX PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX DR
XX N-PSDB; ABL01900.
XX DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT

PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX XX
PS Disclosure; SEQ ID NO 183; 21pp + Sequence Listing; English.
XX XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 512 AA;
Query Match 68.8%; Score 44; DB 22; Length 512;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 HKWHWY 8
DB 221 HPWHWY 226
RESULT 5
AAU14133
ID AAU14133 standard; Protein; 554 AA.
XX AC
XX AAU14133;
XX AC
XX 24-OCT-2001 (first entry)
XX DT
XX Human novel protein #4.
XX DE
XX Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX KW
XX Homo sapiens.
XX OS
XX WO200155437-A2.
XX PN
XX 02-AUG-2001.
XX PD
XX 25-JAN-2001; 2001WO-US02623.
XX PF
XX 25-JAN-2000; 2000US-0491404.
XX PR
XX (HYSE-) HYSEQ INC.
XX PA
XX Tang YT, Liu C, Drmanac RT;
XX PI
XX WPI; 2001-451939/48.
XX DR
XX N-PSDB; AAS22438.
XX DR
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX XX
XX Example 4; Page 522-523; 894pp; English.
XX PS
XX The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC

CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour, in assays to determine biological activity, to
 CC raise antibodies/ elicit an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.

XX SQ Sequence 554 AA;

Query Match 68.8%; Score 44; DB 22; Length 554;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 HKWHWY 8
 | | | | |
 Db 306 HPWHWY 311

RESULT 6

ABB59504
 ID ABB59504 standard; Protein; 683 AA.

XX AC ABB59504;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide' SEQ ID NO 5304.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX FR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL03607.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -

XX PS Disclosure; SEQ ID NO 5304; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 sequences (ABL01840-ABL16175) and the encoded proteins
 (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 683 AA;

Query Match 68.8%; Score 44; DB 22; Length 683;
 Best Local Similarity 82.5%; Pred. No. 1.6e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LKHWHWYV 9
 | | | | |
 Db 207 LHHWHWHL 214

RESULT 7

ABB57998

ID ABB57998 standard; Protein; 684 AA.

XX AC ABB57998;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 786.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL02101.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -

XX PS Disclosure; SEQ ID NO 786; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 sequences (ABL01840-ABL16175) and the encoded proteins
 (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 684 AA;

Query Match 68.8%; Score 44; DB 22; Length 684;
 Best Local Similarity 62.5%; Pred. No. 1.6e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY      2 LHKWHWYV 9
DB      || |||::
        206 LHHWHWHL 213

RESULT 8
AAW1441
ID AAW1441 standard; Protein; 685 AA.
XX
AC AAW1441;
XX
DT 19-JUN-1996 (first entry)
XX
DE Bombyx mori (pro)phenoloxidase Phe52-Gly685.
XX
KW Silkorm; larvae; pro; phenoloxidase; Phe52-Val693; Phe52-Gly685
KW labelling oxidase.
XX
OS Bombyx mori.
XX
FH Key Location/Qualifiers
FT Peptide 6..23
FT Peptide /note= "peptide fragment"
FT Peptide 24..40
FT Peptide /note= "peptide fragment"
FT Peptide 41..51
FT Peptide /note= "peptide fragment"
FT Peptide 52..61
FT Peptide /note= "peptide fragment"
XX
PN JP07289251-A.
XX
PD 07-NOV-1995.
XX
PF 22-APR-1994; 94JP-0085096.
XX
PR 22-APR-1994; 94JP-0085096.
XX
PA (WAKP ) WAKO PURE CHEM IND LTD.
XX
WPI; 1996-015259/02.
DR N-PSDB; AAW1441.
XX
(P)phenol:oxidase from silkworm - useful as a labelling oxidase
Claim 2; Pages 14-17; 29pp; Japanese.
XX
The B. mori silkworm larvae derived (pro)phenoloxidases Phe52-Val693
(AAW1441) and Phe52-Gly685 (AAW1441), are encoded by AAT10240 and
AAT10241, respectively. The (pro)phenoloxidases can be used as new
labelling oxidases.
XX
SQ Sequence 685 AA;
Query Match 68.8%; Score 44; DB 17; Length 685;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 LHKWHWYV 9
DB      || |||::
        207 LHHWHWHL 214

RESULT 9
AAW1441
ID AAW1441 standard; Protein; 685 AA.
XX
AC AAW1441;
XX
DT 15-MAY-1997 (first entry)
XX
PR 15-MAY-1997 (first entry)
XX
DE Prophenol oxidase.

oxidase; prophenol; phenol; domestic silkworm; label; detection.
Bombyx mori.
Key Location/Qualifiers
Peptide 1..52
Protein /label= pro-sequence
53..685 /label= phenol oxidase
/label= "claim 1"
XX
PN JP09023886-A.
XX
PD 28-JAN-1997.
XX
PF 13-JUL-1995; 95JP-0177444.
XX
PR 13-JUL-1995; 95JP-0177444.
XX
PA (WAKP ) WAKO PURE CHEM IND LTD.
XX
WPI; 1997-148588/14.
DR N-PSDB; AAT62660.
XX
(P)phenol oxidase derived from a domestic silkworm - useful as a
labelling oxidase and in pro-phenol oxidase activation system for
detection of microorganisms
Claim 2; Page 14-17; 18pp; Japanese.
XX
This protein is a prophenol oxidase derived from a domestic silkworm.
The phenol oxidase may be used as a novel labelling oxidase. The
elucidation of the primary structure of the prophenol oxidase will
contribute to the reconstitution of a prophenol oxidase activation system
which can be applied to the detection of microorganisms by measurement of
beta-1,3-glucan and peptide glycan.
XX
SQ Sequence 685 AA;
Query Match 68.8%; Score 44; DB 18; Length 685;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 LHKWHWYV 9
DB      || |||::
        207 LHHWHWHL 214

RESULT 10
AAU34057
ID AAU34057 standard; Protein; 142 AA.
XX
AC AAU34057;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #333.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2001; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.

```


23-OCT-2000; 2000US-242578P.
 27-NOV-2000; 2000US-253625P.
 22-DEC-2000; 2000US-257931P.
 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI; 2001-611495/70.
 DR N-PSDB; AAS51916.

XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 5553; 51lpp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Seq Sequence 142 AA;

Query Match 67.2%; Score 43; DB 22; Length 142;
 Best Local Similarity 75.0%; Pred. No. 50;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWYV 9
 || : ||||
 Db 21 LHNFWHYV 28

RESULT 11
 AAU36776
 ID AAU36776 standard; Protein; 147 AA.

XX AAU36776;

DT 14-FEB-2002 (first entry)

DE Staphylococcus aureus cellular proliferation protein #946.

KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.

OS Staphylococcus aureus.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.
 27-NOV-2000; 2000US-253625P.
 22-DEC-2000; 2000US-257931P.
 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI; 2001-611495/70.
 DR N-PSDB; AAS54635.

XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 12369; 51lpp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Seq Sequence 147 AA;

Query Match 67.2%; Score 43; DB 22; Length 147;
 Best Local Similarity 75.0%; Pred. No. 52;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWYV 9
 || : ||||
 Db 25 LHNFWHYV 32

RESULT 12
 ABP39358
 ID ABP39358 standard; Protein; 166 AA.

XX ABP39358;

DT 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4203.

KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy.

OS Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

```
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI: 2002-381255/41.
XX DR N-PSDB; ABN91903.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 4203; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX CC antibacterial activity and can be used in gene therapy. The sequences
XX CC can also be used in the diagnosis and treatment of bacterial infections,
XX CC particularly S. epidermidis infections. The sequences can be used to
XX CC screen for compounds able to interfere with the S. epidermidis life
XX CC cycle or inhibit S. epidermidis infection.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC USPTO web site.
XX
XX Sequence 166 AA;
SQ Query Match 67.2%; Score 43; DB 23; Length 166;
Best Local Similarity 75.0%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LHKWHWYV 9
DB 43 LHNFWYV 50
|||:||||
|||:||||

RESULT 13
ABP26523
ID ABP26523 standard; Protein; 235 AA.
XX
XX ABP26523;
AC
XX
XX 02-JUL-2002 (first entry)
DT
XX
XX Streptococcus polypeptide SEQ ID NO 2222.
DE
XX
XX Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus pyogenes.
OS
XX
XX WO200234771-A2.
PN
XX
XX 02-MAY-2002.
PD
XX
XX 29-OCT-2001; 2001WO-GB04789.
PF
XX
XX 27-OCT-2000; 2000GB-0026333.
PR
XX 24-NOV-2000; 2000GB-0028727.
PR
XX 07-MAR-2001; 2001GB-0005640.
PR
XX
XX (CHIR-) CHIRON SPA.
PA
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX
XX WPI; 2002-352536/38.
XX DR N-PSDB; ABN67154.
XX
XX New Streptococcus protein for the treatment or prevention of infection
XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX PT for detecting a compound that binds to the protein -
XX
```

```
XX
XX Claim 1; Page 3373; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins.
XX
XX Sequence 235 AA;
SQ Query Match 67.2%; Score 43; DB 23; Length 235;
Best Local Similarity 40.0%; Pred. No. 82;
Matches 6; Conservative 3; Mismatches 0; Indels 6; Gaps 1;
QY 1 FLHKW-----HWYV 9
DB 82 FWERWATKGLHWYL 96
|||:||||
|||:||||

RESULT 14
ABG19706
ID ABG19706 standard; Protein; 1248 AA.
XX
XX ABG19706;
AC
XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #19697.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
XX PI
XX
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS83893.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
XX Claim 20; SEQ ID No 50065; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
```

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1248 AA;

Query Match 67.2%; Score 43; DB 22; Length 1248;
Best Local Similarity 83.3%; Pred. No. 4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKWHW 7
||:||||
Db 1065 LQWHW 1070

RESULT 15
ABJ10988
ID ABJ10988 standard; Protein; 98 AA.

AC ABJ10988;

DT 10-DEC-2002 (first entry)

DE Yeast selected interacting domain protein SEQ ID NO: 110.

KW Yeast; protein-protein interaction; Selected Interacting Domain;
KW SID (RTM); secretion yield; cancer; neurodegenerative disease; fungicide;
KW cytostatic; neuroprotective.

OS Saccharomyces cerevisiae.

PN WO200266504-A2.

PD 29-AUG-2002.

PF 14-FEB-2002; 2002WO-EP02299.

PR 16-FEB-2001; 2001US-269266P.

PA (HYBR-) HYBRIGENICS.

PI Legrain P;

XX WPI; 2002-674913/72.

DR N-PSDB; ABT11305.

PT New protein-protein complexes of Saccharomyces cerevisiae, useful in
PT drug screening or development, for developing yeast strains with better
PT secretion yield of protein, or in gene therapy (e.g. to treat Candida
PT infection or cancer)

PS Claim 6; Page 71; 357pp; English.

XX The present invention relates to complexes between Saccharomyces
CC cerevisiae selected interacting domain (SID (RTM)) proteins and coding
CC sequences. The protein complexes of S. cerevisiae are useful in drug
CC development, in screening drugs or agents that modulate the interaction
CC of proteins, for developing yeast strains with better secretion yield of

CC protein, and in gene therapy. The protein complexes, polypeptides and
CC polynucleotides are useful for preventing or treating Candida infection,
CC cancer or neurodegenerative diseases in humans or animals. The present
CC sequence is a protein of the invention.

SQ Sequence 98 AA;

Query Match 65.6%; Score 42; DB 23; Length 98;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
|||||
Db 74 KWHWY 78

Search completed: December 16, 2003, 14:14:29
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:11:00 ; Search time 12.3333 Seconds
(without alignments)
30.875 Million cell updates/sec

Title: US-09-870-089B-9

Perfect score: 64

Sequence: 1 FLHKWHYV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	70.3	191	4	US-09-107-532A-5382
2	43	67.2	166	4	US-09-134-001C-4203
3	42	65.6	480	3	US-09-182-859-4
4	42	65.6	480	3	US-09-170-670-5
5	42	65.6	480	3	US-09-193-068-5
6	42	65.6	480	3	US-09-183-412-5
7	42	65.6	480	4	US-09-290-734-5
8	42	65.6	480	4	US-09-672-459-4
9	42	65.6	480	4	US-09-545-586-5
10	42	65.6	483	1	US-08-700-34
11	42	65.6	483	1	US-08-645-971-2
12	42	65.6	483	2	US-08-468-220-32
13	42	65.6	483	2	US-08-468-220-36
14	42	65.6	483	2	US-08-468-698-32
15	42	65.6	483	2	US-08-468-698-36
16	42	65.6	483	2	US-08-704-706A-34
17	42	65.6	483	2	US-08-600-908A-13
18	42	65.6	483	3	US-08-890-383-3
19	42	65.6	483	3	US-08-683-838A-13
20	42	65.6	483	3	US-08-914-679A-3
21	42	65.6	483	3	US-09-182-859-2
22	42	65.6	483	3	US-09-170-670-4
23	42	65.6	483	3	US-09-193-068-4
24	42	65.6	483	3	US-09-183-412-4
25	42	65.6	483	3	US-08-985-659-35
26	42	65.6	483	3	US-09-264-097-2
27	42	65.6	483	3	US-08-194-664A-32

28 42 65.6 483 3 US-08-194-664A-36 Sequence 36, Appl
29 42 65.6 483 4 US-09-291-023A-16 Sequence 16, Appl
30 42 65.6 483 4 US-09-291-023A-21 Sequence 21, Appl
31 42 65.6 483 4 US-09-290-734-4 Sequence 4, Appl
32 42 65.6 483 4 US-09-537-168-6 Sequence 4, Appl
33 42 65.6 483 4 US-09-537-168-6 Sequence 6, Appl
34 42 65.6 483 4 US-09-672-459-2 Sequence 2, Appl
35 42 65.6 483 4 US-09-636-252A-13 Sequence 13, Appl
36 42 65.6 483 4 US-09-381-687-6 Sequence 6, Appl
37 42 65.6 483 4 US-09-381-687-8 Sequence 8, Appl
38 42 65.6 483 4 US-09-545-586-4 Sequence 4, Appl
39 42 65.6 483 5 PCT-US94-01553A-32 Sequence 32, Appl
40 42 65.6 483 5 PCT-US94-01553A-36 Sequence 36, Appl
41 42 65.6 483 5 PCT-US95-10426-32 Sequence 32, Appl
42 42 65.6 483 5 PCT-US95-10426-36 Sequence 36, Appl
43 42 65.6 487 2 US-08-468-220-37 Sequence 37, Appl
44 42 65.6 487 2 US-08-468-698-37 Sequence 37, Appl
45 42 65.6 487 3 US-08-194-664A-37 Sequence 37, Appl

ALIGNMENTS

RESULT 1

US-09-107-532A-5382
; Sequence 5382 Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5382:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...191
; SEQUENCE DESCRIPTION: SEQ ID NO: 5382:
US-09-107-532A-5382

```
Query Match      70.3%; Score 45; DB 4; Length 191;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKWHWYV 9
Db 52 LHOYHWYV 59

RESULT 2
US-09-134-001C-4203
; Sequence 4203, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4203
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4203

Query Match      67.2%; Score 43; DB 4; Length 166;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWYV 9
Db 43 LHNFWYV 50

RESULT 3
US-09-182-859-4
; Sequence 4, Application US/09182859
; Patent No. 6143708
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben
; APPLICANT: Bisgaard-Frantzen, Henrik
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 4796.204-US
; CURRENT APPLICATION NUMBER: US/09/182,859
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 0515/96
; EARLIER FILING DATE: 1996-04-30
; EARLIER APPLICATION NUMBER: 0712/96
; EARLIER FILING DATE: 1996-06-28
; EARLIER APPLICATION NUMBER: 0775/96
; EARLIER FILING DATE: 1996-07-11
; EARLIER APPLICATION NUMBER: 1263/96
; EARLIER FILING DATE: 1996-11-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-09-182-859-4

Query Match      65.6%; Score 42; DB 3; Length 480;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KWHWY 8
Db 152 KWHWY 156

RESULT 4
US-09-170-670-5
; Sequence 5, Application US/09170670
; Patent No. 6187576
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben
; APPLICANT: Bisgaard-Frantzen, Henrik
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 5276.200-US
; CURRENT APPLICATION NUMBER: US/09/170,670
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 1172/97
; EARLIER FILING DATE: 1997-10-13
; EARLIER APPLICATION NUMBER: 60/063,306
; EARLIER FILING DATE: 1997-10-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 480
; TYPE: PRT
; ORGANISM: B. amyloliquefaciens
US-09-170-670-5

Query Match      65.6%; Score 42; DB 3; Length 480;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KWHWY 8
Db 152 KWHWY 156

RESULT 5
US-09-193-068-5
; Sequence 5, Application US/09193068
; Patent No. 6197565
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Kjulf, S ren
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Andersen, Carsten
; TITLE OF INVENTION: -Amylase Variants
; FILE REFERENCE: 5709.000-US
; CURRENT APPLICATION NUMBER: US/09/193,068
; CURRENT FILING DATE: 1998-11-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 480
; TYPE: PRT
; ORGANISM: B. amyloliquefaciens
US-09-193-068-5

Query Match      65.6%; Score 42; DB 3; Length 480;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KWHWY 8
Db 152 KWHWY 156

RESULT 6
US-09-183-412-5
; Sequence 5, Application US/09183412
; Patent No. 6204232
; GENERAL INFORMATION:
```

```

; APPLICANT: Borchert, Torben V.
; APPLICANT: Svendsen, Allan
; APPLICANT: Andersen, Carsten
; APPLICANT: Nielsen, Bjarne
; APPLICANT: Nielsen, Torben L.
; APPLICANT: Kiaerulff, Soren
; TITLE OF INVENTION: Alpha-Amulase Mutants
; FILE REFERENCE: 5368-200-US
; CURRENT APPLICATION NUMBER: US/09/183,412
; CURRENT FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: 60/064,662
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 60/093,234
; EARLIER FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 1240/97
; EARLIER FILING DATE: 1997-10-30
; EARLIER APPLICATION NUMBER: PA 1998 00936
; EARLIER FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Bacillus amyloliquifaciens
US-09-183-412-5

```

```

Query Match          65.6%; Score 42; DB 3; Length 480;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      4 KWHWY 8
Db      152 KHWY 156

```

```

RESULT 7
US-09-290-734-5
; Sequence 5, Application US/09290734
; Patent No. 6361989
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Bisgard-Frantzen Henrik
; APPLICANT: Outtrup, Helle
; APPLICANT: Nielsen, Bjarne Ronfeldt
; APPLICANT: Nielsen, Vibeke Skovgaard
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: No. 6361989el -Amylase And -Amylase Mutants
; FILE REFERENCE: 5276.400-US
; CURRENT APPLICATION NUMBER: US/09/290,734
; CURRENT FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 480
; TYPE: PRT
; ORGANISM: B. amyloliquifaciens
US-09-290-734-5

```

```

Query Match          65.6%; Score 42; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      4 KWHWY 8
Db      152 KHWY 156

```

```

RESULT 8
US-09-672-459-4
; Sequence 4, Application US/09672459
; Patent No. 6436888
; GENERAL INFORMATION:

```

```

; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben
; APPLICANT: Bisgard-Frantzen, Henrik
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 4796.204-US
; CURRENT APPLICATION NUMBER: US/09/672,459
; CURRENT FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 09/182,859
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 0515/96
; PRIOR FILING DATE: 1996-04-30
; PRIOR APPLICATION NUMBER: 0712/96
; PRIOR FILING DATE: 1996-06-28
; PRIOR APPLICATION NUMBER: 0775/96
; PRIOR FILING DATE: 1996-07-11
; PRIOR APPLICATION NUMBER: 1263/96
; PRIOR FILING DATE: 1996-11-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Bacillus amyloliquifaciens
US-09-672-459-4

```

```

Query Match          65.6%; Score 42; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      4 KWHWY 8
Db      152 KHWY 156

```

```

RESULT 9
US-09-545-586-5
; Sequence 5, Application US/09545586
; Patent No. 6528298
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Bisgard-Frantzen Henrik
; APPLICANT: Outtrup, Helle
; APPLICANT: Nielsen, Bjarne Ronfeldt
; APPLICANT: Nielsen, Vibeke Skovgaard
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: No. 6528298el -Amylase And -Amylase Mutants
; FILE REFERENCE: 5276.400-US
; CURRENT APPLICATION NUMBER: US/09/545,586
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: US/09/290,734
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 480
; TYPE: PRT
; ORGANISM: B. amyloliquifaciens
US-09-545-586-5

```

```

Query Match          65.6%; Score 42; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      4 KWHWY 8
Db      152 KHWY 156

```

```

RESULT 10
US-08-468-700-34
; Sequence 34, Application US/08468700
; Patent No. 5736499

```

GENERAL INFORMATION:
APPLICANT: COLIN MITCHINSON
APPLICANT: CAROL A. REQUADT
APPLICANT: TRACI H. ROPP
APPLICANT: LEIF P. SOLHEIM
TITLE OF INVENTION: MUTANT ALPHA-AMYLASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,700
FILING DATE: 06-JUN-1995
CLASSIFICATION: 252
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC275
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 483 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-468-700-34

Query Match 65.6%; Score 42; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KHWY 8
Db 154 KHWY 158
|||||

RESULT 11
US-08-645-971-2
Sequence 2, Application US/08645971
Patent No. 5763385
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Modified Alpha-Amylases Having Altered
TITLE OF INVENTION: Calcium Binding Properties
NUMBER OF SEQUENCES: 5
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA: US/08/645,971
APPLICATION NUMBER: US/08/645,971
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 483 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacillus licheniformis

US-08-645-971-2
Query Match 65.6%; Score 42; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KHWY 8
Db 154 KHWY 158
|||||

RESULT 12
US-08-468-220-32
Sequence 32, Application US/08468220
Patent No. 5824532
GENERAL INFORMATION:
APPLICANT: Antrim, Richard L.
APPLICANT: Barnett, Christopher
APPLICANT: Mitchinson, Colin
APPLICANT: Power, Scott D.
APPLICANT: Requaadt, Carol
APPLICANT: Solheim, Leif P.
TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,220
FILING DATE: 06-JUN-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/194,664
FILING DATE: 10-FEB-94
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/016,395
FILING DATE: 11-FEB-93
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher L.
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC220D1
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 483 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-468-220-32

Query Match 65.6%; Score 42; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KHWY 8
Db 154 KHWY 158
|||||

RESULT 13
US-08-468-220-36

; Sequence 36, Application US/08468220
; Patent No. 5824532
; GENERAL INFORMATION:
; APPLICANT: Antrim, Richard L.
; APPLICANT: Barnett, Christopher
; APPLICANT: Mitchinson, Colin
; APPLICANT: Power, Scott D.
; APPLICANT: Reguadt, Carol
; APPLICANT: Solheim, Leif P.
; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/468,220
; FILING DATE: 06-JUN-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/194,664
; FILING DATE: 10-FEB-94
; APPLICATION NUMBER: 08/016,395
; FILING DATE: 11-FEB-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC220D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7217
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-220-36

Query Match 65.6%; Score 42; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KHWY 8
Db 154 KHWY 158

RESULT 14
US-08-468-698-32
; Sequence 32, Application US/08468698
; Patent No. 5849549
; GENERAL INFORMATION:
; APPLICANT: Antrim, Richard L.
; APPLICANT: Barnett, Christopher
; APPLICANT: Mitchinson, Colin
; APPLICANT: Power, Scott D.
; APPLICANT: Reguadt, Carol
; APPLICANT: Solheim, Leif P.
; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.

; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/468,698
; FILING DATE: 06-JUN-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/194,664
; FILING DATE: 10-FEB-94
; APPLICATION NUMBER: 08/016,395
; FILING DATE: 11-FEB-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC220D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7217
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-698-32

Query Match 65.8%; Score 42; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KHWY 8
Db 154 KHWY 158

RESULT 15
US-08-468-698-36
; Sequence 36, Application US/08468698
; Patent No. 5849549
; GENERAL INFORMATION:
; APPLICANT: Antrim, Richard L.
; APPLICANT: Barnett, Christopher
; APPLICANT: Mitchinson, Colin
; APPLICANT: Power, Scott D.
; APPLICANT: Reguadt, Carol
; APPLICANT: Solheim, Leif P.
; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/468,698
; FILING DATE: 06-JUN-95


```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/194,664
; FILING DATE: 10-FEB-94
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/016,395
; FILING DATE: 11-FEB-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC220D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7555
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-698-36
```

```
Query Match 65.6%; Score 42; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 4 KWHWY 8
Db 154 KWHWY 158
```

```
Search completed: December 16, 2003, 14:20:40
Job time : 12.3333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:18:11 ; Search time 22.1667 Seconds
(without alignments)
75.512 Million cell updates/sec

Title: US-09-870-089B-9
Perfect score: 64
Sequence: 1 FLKWHWYV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	9	12	US-09-870-089B-9
2	56	87.5	9	12	US-09-870-089B-7
3	43	67.2	142	9	US-09-815-242-5553
4	43	67.2	147	9	US-09-815-242-12369
5	42	65.6	480	9	US-09-769-864-5
6	42	65.6	480	12	US-10-186-042-4
7	42	65.6	483	9	US-09-769-864-4
8	42	65.6	483	9	US-09-854-346-8
9	42	65.6	483	9	US-09-854-346-10
10	42	65.6	483	10	US-09-918-543-8
11	42	65.6	483	10	US-09-918-543-10
12	42	65.6	483	10	US-09-918-543-30
13	42	65.6	483	12	US-09-925-576C-8
14	42	65.6	483	12	US-09-925-576C-10
15	42	65.6	483	12	US-10-184-771-13

16	42	65.6	483	12	US-10-186-042-2	Sequence 2, Appli
17	42	65.6	483	15	US-10-146-327-4	Sequence 4, Appli
18	42	65.6	483	15	US-10-146-327-6	Sequence 6, Appli
19	42	65.6	512	12	US-10-105-733-8	Sequence 8, Appli
20	42	65.6	512	12	US-10-081-739A-8	Sequence 8, Appli
21	42	65.6	512	12	US-10-184-771-2	Sequence 2, Appli
22	42	65.6	512	12	US-10-199-922-2	Sequence 2, Appli
23	42	65.6	512	15	US-10-081-872-114	Sequence 114, App
24	42	65.6	514	12	US-10-184-771-4	Sequence 4, Appli
25	41	64.1	9	12	US-09-870-089B-3	Sequence 3, Appli
26	41	64.1	11	9	US-09-780-070-21	Sequence 21, Appli
27	41	64.1	449	9	US-09-732-618-24	Sequence 24, Appli
28	41	64.1	449	9	US-09-732-618-25	Sequence 25, Appli
29	40	62.5	9	12	US-09-870-089B-5	Sequence 5, Appli
30	40	62.5	374	15	US-10-225-567A-302	Sequence 302, App
31	40	62.5	412	15	US-10-156-761-9146	Sequence 9146, Ap
32	39	60.9	153	15	US-10-156-761-7672	Sequence 7672, Ap
33	38	59.4	46	15	US-10-106-698-5214	Sequence 5214, Ap
34	38	59.4	130	12	US-10-238-075-693	Sequence 693, App
35	38	59.4	552	10	US-09-843-676-4	Sequence 4, Appli
36	38	59.4	552	10	US-09-766-253-4	Sequence 4, Appli
37	38	59.4	552	11	US-09-438-486-4	Sequence 4, Appli
38	38	59.4	552	15	US-10-053-758-4	Sequence 4, Appli
39	38	59.4	552	15	US-10-054-295-4	Sequence 4, Appli
40	38	59.4	552	15	US-10-054-611-4	Sequence 4, Appli
41	37	57.8	19	12	US-10-029-386-33052	Sequence 33052, A
42	37	57.8	42	9	US-09-864-761-34285	Sequence 34285, A
43	37	57.8	61	15	US-10-076-604-204	Sequence 204, App
44	37	57.8	139	9	US-09-864-761-33721	Sequence 33721, A
45	37	57.8	178	9	US-09-815-242-13494	Sequence 13494, A

ALIGNMENTS

RESULT 1
US-09-870-089B-9
; Sequence 9, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870,089B
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATF4/CREB-2
US-09-870-089B-9

Query Match 100.0%; Score 64; DB 12; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FLKWHWYV 9
Db 1 FLKWHWYV 9

RESULT 2
US-09-870-089B-7
; Sequence 7, Application US/09870089B
; Publication No. US20030175252A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Nicolette
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS FOR OVARIAN CANCER
; FILE REFERENCE: 68126881209900
; CURRENT APPLICATION NUMBER: US/09/870,089B

; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ATP4/CREB-2
US-09-870-089B-7

Query Match 87.5%; Score 56; DB 12; Length 9;
Best Local Similarity 88.9%; Pred. No. 6.1e+05;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLKWHWYV 9
||| |||||
Db 1 FLKWHWYV 9

RESULT 3
US-09-815-242-5553

; Sequence 5553, Application US/09815242
; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5553

; LENGTH: 142

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-5553

Query Match 67.2%; Score 43; DB 9; Length 142;
Best Local Similarity 75.0%; Pred. No. 73;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWYV 9
||| :|||
Db 21 LHNFWYV 28

RESULT 4
US-09-815-242-12369
; Sequence 12369, Application US/09815242
; Patent No. US20020061569A1

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12369

; LENGTH: 147

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12369

Query Match 67.2%; Score 43; DB 9; Length 147;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWYV 9
||| :|||
Db 25 LHNFWYV 32

RESULT 5

US-09-769-864-5

; Sequence 5, Application US/09769864

; Patent No. US20010039253A1

; GENERAL INFORMATION:

; APPLICANT: Borchert, Torben V.

; APPLICANT: Svendsen, Allan

; APPLICANT: Andersen, Carsten

; APPLICANT: Nielsen, Bjarne

; APPLICANT: Nielsen, Torben L.

; APPLICANT: Kjaerulff, Soren

; TITLE OF INVENTION: Alpha-Amulase Mutants

; FILE REFERENCE: 5368.200-US

; CURRENT APPLICATION NUMBER: US/09/769,864

; CURRENT FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 09/183,412

; PRIOR FILING DATE: 1998-10-30

; NUMBER OF SEQ ID NOS: 58

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 5

; LENGTH: 480

; TYPE: PRT

; ORGANISM: Bacillus amyloquifaciens

US-09-769-864-5

Query Match 65.6%; Score 42; DB 9; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
|||||

Db 152 KWHWY 156

RESULT 6

US-10-186-042-4
; Sequence 4, Application US/10186042
; Publication No. US20030171236A1
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben
; APPLICANT: Bisgaard-Prantzen, Henrik
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 4796.204-US
; CURRENT APPLICATION NUMBER: US/10/186,042
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/672,459
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 09/182,859
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 0515/96
; PRIOR FILING DATE: 1996-04-30
; PRIOR APPLICATION NUMBER: 0712/96
; PRIOR FILING DATE: 1996-06-28
; PRIOR APPLICATION NUMBER: 0775/96
; PRIOR FILING DATE: 1996-07-11
; PRIOR APPLICATION NUMBER: 1263/96
; PRIOR FILING DATE: 1996-11-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-10-186-042-4

Query Match 65.6%; Score 42; DB 12; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
|||||

Db 152 KWHWY 156

RESULT 7

US-09-769-864-4
; Sequence 4, Application US/09769864
; Patent No. US20010039253A1
; GENERAL INFORMATION:
; APPLICANT: Borchert, Torben V.
; APPLICANT: Svendsen, Allan
; APPLICANT: Andersen, Carsten
; APPLICANT: Nielsen, Bjarne
; APPLICANT: Nissen, Torben L.
; APPLICANT: Kjaerulff, Soren
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 5368.200-US
; CURRENT APPLICATION NUMBER: US/09/769,864
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/183,412
; PRIOR FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-769-864-4

Query Match 65.6%; Score 42; DB 9; Length 483;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
|||||

Db 154 KWHWY 158

RESULT 8

US-09-854-346-8
; Sequence 8, Application US/09854346
; Patent No. US20020068352A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020068352A1ozymes A/S
; APPLICANT: Svendsen, Allan
; APPLICANT: Jorgensen, Christel Thea
; APPLICANT: Nielsen, Bjarne Ronfeldt
; TITLE OF INVENTION: Alpha-amylase variants with altered 1,6 activity
; FILE REFERENCE: 6140.200-US
; CURRENT APPLICATION NUMBER: US/09/854,346
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-854-346-8

Query Match 65.6%; Score 42; DB 9; Length 483;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
|||||

Db 154 KWHWY 158

RESULT 9

US-09-854-346-10
; Sequence 10, Application US/09854346
; Patent No. US20020068352A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020068352A1ozymes A/S
; APPLICANT: Svendsen, Allan
; APPLICANT: Jorgensen, Christel Thea
; APPLICANT: Nielsen, Bjarne Ronfeldt
; TITLE OF INVENTION: Alpha-amylase variants with altered 1,6 activity
; FILE REFERENCE: 6140.200-US
; CURRENT APPLICATION NUMBER: US/09/854,346
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-09-854-346-10

Query Match 65.6%; Score 42; DB 9; Length 483;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
|||||

Db 152 KWHWY 156

RESULT 10

US-09-918-543-8
; Sequence 8, Application US/09918543
; Patent No. US20020155574A1
; GENERAL INFORMATION:

```
; APPLICANT: No. US20020155574Allozymes A/S
; APPLICANT: Thisted, Thomas
; APPLICANT: Kjaerulff, Soren
; APPLICANT: Andersen, Carsten
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Alpha-amylase mutants with altered properties
; FILE REFERENCE: 10062.200-US
; CURRENT APPLICATION NUMBER: US/09/918,543
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-918-543-8

Query Match      65.6%; Score 42; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 KWHWY 8
Db      154 KWHWY 158

RESULT 11
US-09-918-543-10
; Sequence 10, Application US/09918543
; Patent No. US20020155574A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020155574Allozymes A/S
; APPLICANT: Thisted, Thomas
; APPLICANT: Kjaerulff, Soren
; APPLICANT: Andersen, Carsten
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Alpha-amylase mutants with altered properties
; FILE REFERENCE: 10062.200-US
; CURRENT APPLICATION NUMBER: US/09/918,543
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-09-918-543-10

Query Match      65.6%; Score 42; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 KWHWY 8
Db      152 KWHWY 156

RESULT 12
US-09-918-543-30
; Sequence 30, Application US/09918543
; Patent No. US20020155574A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020155574Allozymes A/S
; APPLICANT: Thisted, Thomas
; APPLICANT: Kjaerulff, Soren
; APPLICANT: Andersen, Carsten
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Alpha-amylase mutants with altered properties
; FILE REFERENCE: 10062.200-US
; CURRENT APPLICATION NUMBER: US/09/918,543
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-918-543-30

Query Match      65.6%; Score 42; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 KWHWY 8
Db      154 KWHWY 158

RESULT 13
US-09-925-576C-8
; Sequence 8, Application US/09925576C
; Publication No. US20030129718A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Carsten
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Nielsen, Bjarne Ronfeldt
; TITLE OF INVENTION: Amylase Variants
; FILE REFERENCE: 10004.204-US
; CURRENT APPLICATION NUMBER: US/09/925,576C
; CURRENT FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-925-576C-8

Query Match      65.6%; Score 42; DB 12; Length 483;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 KWHWY 8
Db      154 KWHWY 158

RESULT 14
US-09-925-576C-10
; Sequence 10, Application US/09925576C
; Publication No. US20030129718A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Carsten
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Nielsen, Bjarne Ronfeldt
; TITLE OF INVENTION: Amylase Variants
; FILE REFERENCE: 10004.204-US
; CURRENT APPLICATION NUMBER: US/09/925,576C
; CURRENT FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-09-925-576C-10

Query Match      65.6%; Score 42; DB 12; Length 483;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 KWHWY 8
Db      152 KWHWY 156
```

```
RESULT 15
US-10-184-771-13
; Sequence 13, Application US/10184771
; Publication No. US20030170769A1
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Borchert, Torben Vedel
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 0776/1F216-US2
; CURRENT APPLICATION NUMBER: US/10/184,771
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/636,252
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 08/683,838
; PRIOR FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bascillius
US-10-184-771-13

Query Match      65.6%; Score 42; DB 12; Length 483;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 KHWY 8
Db      152 KHWY 156
```

Search completed: December 16, 2003, 14:41:06
Job time : 22.1667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:10:15 / Search time 10.3333 Seconds
(without alignments)
83.760 Million cell updates/sec

Title: US-09-870-089b-9

Perfect score: 64

Sequence: 1 FLKHWHYV 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76.*

1: Piri.*

2: Piri.*

3: Piri.*

4: Piri.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	45	70.3	153	2 G69660	metal-regulated pr
2	44	68.8	143	2 H69618	stress- and starva
3	44	68.8	159	2 C75311	conserved hypothet
4	44	68.8	498	2 T51430	dolichyl-phosphate
5	44	68.8	554	2 S71751	dolichyl-phosphate
6	43	67.2	147	2 H90007	general stress pro
7	43	67.2	209	2 H90007	conserved hypothet
8	43	67.2	218	2 G82786	conserved hypothet
9	43	67.2	239	2 F70060	N-acetylphosphinot
10	43	67.2	299	2 A47031	biolaphos acetylhy
11	43	67.2	631	1 BHTLA	hemocyanin chain a
12	42	65.6	283	2 T16945	hypothetical prote
13	42	65.6	512	1 ALBSL	alpha-amylase (EC
14	42	65.6	514	1 ALBSN	alpha-amylase (EC
15	42	65.6	624	1 BHTLE	hemocyanin chain e
16	42	65.6	626	2 S67364	hemocyanin chain e
17	41	64.1	293	2 A11904	hypothetical prote
18	41	64.1	259	2 AG2462	hypothetical prote
19	41	64.1	317	2 AF2129	hypothetical prote
20	41	64.1	493	2 G88979	protein P3784.5 li
21	40	62.5	146	2 B83777	stress- and starva
22	40	62.5	172	2 S28082	vif protein - simi
23	40	62.5	260	2 G75174	hypothetical prote
24	40	62.5	295	2 H71039	hypothetical prote
25	40	62.5	627	1 BHTLD	hemocyanin chain d
26	40	62.5	628	1 BHC2A	hemocyanin subunit
27	39	60.9	274	2 T50986	hypothetical prote
28	39	60.9	393	2 C89861	hypothetical prote
29	39	60.9	538	2 H87347	hypothetical prote

30	39	60.9	544	2 S61985	ALG6 protein - yea
31	39	60.9	582	2 F95871	hypothetical prote
32	39	60.9	611	2 S61147	TCM10 protein - ye
33	39	60.9	989	2 T15576	hypothetical prote
34	38	59.4	145	2 E90791	transposase [impor
35	38	59.4	145	2 A85652	unknown in ISIN [1
36	38	59.4	177	2 JN0097	bacterioferrin F1
37	38	59.4	193	2 JN0846	enterohemolysin 2
38	38	59.4	200	2 T37187	hypothetical prote
39	38	59.4	257	2 A97775	hypothetical prote
40	38	59.4	257	2 D85637	spore coat polysac
41	38	59.4	339	2 S19439	probable membrane
42	38	59.4	512	2 S19439	phosphodiesterase/
43	38	59.4	530	2 AC2085	probable exported
44	38	59.4	569	2 AB0005	hypothetical prote
45	38	59.4	634	2 T22351	hypothetical prote

ALIGNMENTS

RESULT 1

G69660
metal-regulated protein mrgA - Bacillus subtilis
N:Alternate names: metalloregulation DNA-binding stress protein mrgA
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: G69660; S70240; B48656; D44519; S35938
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Buisson, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerich, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaout, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetall
Krieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serc
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terptrak, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G69660
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-153 <KUN>
A:Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15288.1; PID:g2635795
A:Experimental source: strain 168
R:Chen, L.; Helmann, J.D.
Mol. Microbiol. 18, 295-300, 1995
A:Title: Bacillus subtilis MrgA is a Dps(PexB) homologue: evidence for metalloregulatio
A:Reference number: S70240; MUID:96296451; PMID:8709848
A:Accession: S70240
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown
A:Residues: 1-153 <CHB>
A:Cross-references: EMBL:L19547; NID:g852080; PIDN:AAA68042.1; PID:g852076
R:Chen, L.; James, L.P.; Helmann, J.D.
J. Bacteriol. 175, 5428-5437, 1993
A:Title: Metalloregulation in Bacillus subtilis: isolation and characterization of two
A:Reference number: A48656; MUID:93374837; PMID:8396117
A:Accession: B48656
A:Molecule type: DNA
A:Residues: 1-60, 'WIPSLSR' <CH2>
A:Cross-references: EMBL:Z22928; NID:g396219; PIDN:CAA80510.1; PID:g396221
R:Dowds, B.C.A.
submitted to the Protein Sequence Database, February 1993
A:Reference number: A44519
A:Accession: D44519
A:Molecule type: protein
A:Residues: 'X', 2-15, 'KST', 19-20, 'TVFRMH' <DOW>

A;Note: protein was found to over-accumulate in hydrogen peroxide-resistant mutant strain
 C;Genetics:
 A;Gene: mrGA
 C;Superfamily: hypothetical protein H11349

Query Match 70.3%; Score 45; DB 2; Length 153;
 Best Local Similarity 75.0%; Pred. No. 5.5;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHKWHWYV 9
 ||:||||
 Db 31 LHRFWYV 38

RESULT 2
 H69618
 stress- and starvation-induced gene controlled by sigma-B dps - Bacillus subtilis
 C;Species: Bacillus subtilis
 C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C;Accession: H69618
 R;Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleg
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akouchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A;Authors: Yoshikawa, H.P.; Zumstorf, B.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A;Reference number: AG9580; MUID:198044033; PMID:9384377
 A;Accession: H69618
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-145 <KUN>
 A;Cross-references: GB:Z99119; GB:AL009126; NID:G2635411; PIDN:CAB15043.1; PID:G2635549
 A;Experimental source: strain 168
 C;Genetics:
 A;Gene: dps
 C;Superfamily: hypothetical protein H11349

Query Match 68.8%; Score 44; DB 2; Length 145;
 Best Local Similarity 75.0%; Pred. No. 7.4;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWYV 9
 ||:||||
 Db 23 LHNHYWYV 30

RESULT 3
 C5311
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C;Accession: C5311
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: C75311
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-159 <WHI>
 A;Cross-references: GB:AE002048; GB:AE000513; NID:G6459929; PIDN:AAF11686.1; PID:G645993
 A;Experimental source: strain R1

C;Genetics:
 A;Gene: DR2142
 A;Map position: 1

Query Match 68.8%; Score 44; DB 2; Length 159;
 Best Local Similarity 62.5%; Pred. No. 8.1;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWYV 9
 ||:||||
 Db 11 VHPWHWYV 18

RESULT 4
 T51430
 dolichyl-phosphate-mannose-glycolipid alpha-mannosyltransferase-like protein - Arabidops
 N;Alternate names: protein T9L3_150
 C;Species: Arabidopsis thaliana (mouse-ear cross)
 C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
 C;Accession: T51430
 R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
 submitted to the Protein Sequence Database, August 2000
 A;Reference number: Z25394
 A;Accession: T51430
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-498 <SAT>
 A;Cross-references: EMBL:AL391149
 A;Experimental source: cultivar Columbia; BAC clone T9L3
 C;Genetics:
 A;Map position: 5
 A;Introns: 78/2; 117/3; 152/3; 224/2; 318/2; 390/3; 441/2
 A;Note: T9L3_150

Query Match 68.8%; Score 44; DB 2; Length 498;
 Best Local Similarity 83.3%; Pred. No. 25;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWY 8
 ||:||||
 Db 265 HPWHWY 270

RESULT 5
 S71751
 dolichyl-phosphate-mannose-glycolipid alpha-mannosyltransferase (EC 2.4.1.130) PIGB - hu
 N;Alternate names: membrane protein PIG-B; phosphatidyl-inositol glycolipid biosynthesi
 C;Species: Homo sapiens (man)
 C;Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 03-Jun-2002
 C;Accession: S71751
 R;Takahashi, M.; Inoue, N.; Ohishi, K.; Maeda, Y.; Nakamura, N.; Endo, Y.; Fujita, T.; I
 EMBO J. 15, 4254-4261, 1996
 A;Title: PIG-B, a membrane protein of the endoplasmic reticulum with a large luminal don
 A;Reference number: S71751; MUID:97015126; PMID:8861954
 A;Accession: S71751
 A;Molecule type: mRNA
 A;Residues: 1-554 <TAK>
 A;Cross-references: EMBL:D42138; NID:G1552168; PIDN:BAA07709.1; PID:G1552169
 A;Experimental source: cell line P39
 A;Note: part of the genomic DNA was also sequenced
 C;Genetics:
 A;Gene: GDB:PIGB
 A;Cross-references: GDB:9956843; OMIM:604122
 A;Map position: 15q21-15q22
 C;Function:
 A;Description: involved in surface protein binding to the membrane via glycosyl-phosphat
 A;Pathway: GPI-anchor biosynthesis
 C;Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein
 P;60-77/Domain: transmembrane #status predicted <TPM>

Query Match 68.8%; Score 44; DB 2; Length 554;
 Best Local Similarity 83.3%; Pred. No. 28;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 HKWHWY 8
| | | | |
Db 306 HPWHWY 311

RESULT 6

H90007
General stress protein 20U [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: H90007
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A99758; MUID:21311952; PMID:11418146
A:Accession: H90007
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-147 <KUR>
A:Cross-references: GB:BA000018; PID:g13701933; PIDN:BA043225.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: dps
C:Superfamily: hypothetical protein H11349

Query Match 67.2%; Score 43; DB 2; Length 147;
Best Local Similarity 75.0%; Pred. No. 11;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWYV 9
| | : | | | |
Db 25 LHNFWWYV 32

RESULT 7

A87628
conserved hypothetical protein CC3059 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: A87628
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, J.; Armolava, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87628
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <STO>
A:Cross-references: GB:AR005673; NID:g13424707; PIDN:AAK25021.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3059
C:Superfamily: Deinococcus radiodurans hypothetical protein DR1792

Query Match 67.2%; Score 43; DB 2; Length 209;
Best Local Similarity 83.3%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKWHWY 7
| | | | |
Db 17 LQWHWY 22

RESULT 8

G82786
conserved hypothetical protein XF0597 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: G82786

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82786
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-218 <SIM>
A:Cross-references: GB:AE003305; GB:AE003849; NID:g9105456; PIDN:AAF83407.1; GSPDB:GN00149
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, A.; Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Fro, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Lai, Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawas, A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva, M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0597
C:Superfamily: Deinococcus radiodurans hypothetical protein DR1792

Query Match 67.2%; Score 43; DB 2; Length 218;
Best Local Similarity 83.3%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LHKWHWY 7
| | : | | | |
Db 28 LQWHWY 33

RESULT 9

PT0060
N-acetylphosphothricin-tripetide-deacetylase - Streptomyces viridochromogenes
C:Species: Streptomyces viridochromogenes
C:Date: 31-Mar-1992 #sequence_revision 26-Apr-1996 #text_change 19-May-2000
C:Accession: S20686; PT0060
R:Alijah, R.; Hilleman, D.; Nussbaumer, B.; Pelzer, S.; Wohlleben, W.
submitted to the EMBL Data Library, March 1992
A:Description: Gene disruption and gene replacement analysis of a 4 kb BamHI fragment
A:Reference number: S20683
A:Accession: S20686
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <ALI>
A:Cross-references: EMBL:X65195; NID:g47997; PIDN:CAA46315.1; PID:g48001
R:Wohlleben, W.; Arnold, W.; Broer, I.; Hilleman, D.; Strauch, E.; Fuehler, A.
Gene 70, 25-37, 1988
A:Title: Nucleotide sequence of the phosphothricin N-acetyltransferase gene from Str
A:Reference number: JT0409; MUID:89196914; PMID:3240868
A:Accession: PT0060
A:Molecule type: DNA
A:Residues: 1-164 <MOH>
A:Cross-references: GB:M22827; NID:g295177; PIDN:AAAT27110.1; PID:g295180
A:Experimental source: strain Tue 494
C:Superfamily: probable lipolytic protein ybaC

Query Match 67.2%; Score 43; DB 2; Length 299;
Best Local Similarity 83.3%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 HKWHWY 8
| | | | |
Db 267 HWWHWY 272

RESULT 10

A47031
 biaphos acetylhydrolase - Streptomyces hygroscopicus
 C;Species: Streptomyces hygroscopicus
 C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-2000
 C;Accession: A47031
 R;Raibaud, A.; Zalacain, M.; Holt, T.G.; Tizard, R.; Thompson, C.J.
 J. Bacteriol. 173, 4454-4463, 1991
 A;Title: Nucleotide sequence analysis reveals linked N-acetyl hydrolase, thioesterase,

A;Reference number: A47031; MUID:91294191; PMID:2066341
 A;Accession: A47031
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-299 <RAI>
 C;Superfamily: probable lipolytic protein ybaC

Query Match 67.2%; Score 43; DB 2; Length 299;
 Best Local Similarity 83.3%; Pred. No. 21;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWY 8
 | | | | |
 Db 267 HVWHWY 272

RESULT 11

BHTLA
 hemocyanin chain a - American tarantula (Eurypelma californica)
 C;Species: Eurypelma californica
 C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 01-Aug-1997
 C;Accession: A37975; S10606; S08108; S13440
 R;Voit, R.; Feldmaier-Fuchs, G.
 J. Biol. Chem. 265, 19447-19452, 1990
 A;Title: Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs encoding the

A;Reference number: A37975; MUID:91060544; PMID:2246235
 A;Accession: A37975
 A;Molecule type: mRNA
 A;Residues: 1-631 <VOI>
 A;Cross-references: EMBL:X16893
 A;Note: 29-Ala was also found
 R;Schartau, W.; Metzger, W.; Sonner, P.; Geisert, H.; Storz, H.
 Biol. Chem. Hoppe-Seyler 371, 557-565, 1990
 A;Title: Hemocyanins in spiders, XXIII. Complete amino-acid sequence of subunit a of Eur

A;Reference number: S10606; MUID:91025623; PMID:2222854
 A;Accession: S10606
 A;Molecule type: protein
 A;Residues: 2-28, 'A', 30-44, 'D', 46-132, 'I', 134-148, 'I', 150-154, 'L', 156-262, 'M', 264-411, 'I'
 C;Superfamily: hemocyanin
 C;Keywords: copper; hemolymph; oxygen carrier
 F;2-631/Product: hemocyanin chain a #status experimental <MAT>

F;2-156/Domain: 1 <DOI>
 F;157-380/Domain: 2 <DO2>
 F;381-631/Domain: 3 <DO3>
 F;175,179,206/Binding site: copper (His) #status predicted
 F;326,330,366/Binding site: copper (His) #status predicted

Query Match 67.2%; Score 43; DB 1; Length 631;
 Best Local Similarity 71.4%; Pred. No. 44;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWY 9
 | | | | |
 Db 174 HHWHWV 180

RESULT 12

T16945
 hypothetical protein T27F7.3 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C;Accession: T16945

R;Hallsworth, K.
 submitted to the EMBL Data Library, May 1996
 A;Description: The sequence of C. elegans cosmid T27F7.
 A;Reference number: Z18613

A;Accession: T16945
 A;Status: preliminary; translated from GB/EMBL/DBDJ
 A;Molecule type: DNA
 A;Residues: 1-283 <HAL>
 A;Cross-references: EMBL:U58762; NID:gl330398; PID:gl330400; PID:AA00719.1; GSPDB:GNO

A;Experimental source: strain Bristol N2; clone T27F7
 C;Genetics:
 A;Gene: CBSP:T27F7.3
 A;Map position: 2
 A;Introns: 11/1, 45/1, 63/1, 130/3; 248/3

Query Match 65.6%; Score 42; DB 2; Length 283;
 Best Local Similarity 50.0%; Pred. No. 28;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWY 9
 | | | | |
 Db 143 VHPWYWI 150

RESULT 13

ALBSL

alpha-amylase (EC 3.2.1.1) precursor [validated] - Bacillus licheniformis
 N;Alternate names: 1,4-alpha-D-glucan glucanohydrolase
 C;Species: Bacillus licheniformis

C;Date: 30-Jun-1987 #sequence_revision 24-Apr-1998 #text_change 15-Sep-2000
 C;Accession: A91997; B24549; A91796; A21663; I39774; I39772; A26151; S53788; A00844
 R;Yuuki, T.; Nomura, T.; Tezuka, H.; Tsuboi, A.; Yamagata, H.; Tsukagoshi, N.; Udaka, S.
 J. Biochem. 98, 1147-1156, 1985
 A;Title: Complete nucleotide sequence of a gene coding for heat- and pH-stable alpha-am

ase deduced from the DNA sequence.
 A;Reference number: A91997; MUID:86111694; PMID:2418011
 A;Accession: A91997

A;Molecule type: DNA
 A;Residues: 1-162, 'R', 164-512 <YU>

A;Cross-references: GB:X03236; NID:939551; PIDN:CAA26981.1; PID:g39552
 A;Experimental source: ATCC 27811
 R;Gray, G.L.; Mainzer, S.E.; Rey, M.W.; Lamsa, M.H.; Kindle, K.L.; Carmona, C.; Requadt

J. Bacteriol. 166, 635-643, 1986
 A;Title: Structural genes encoding the thermophilic alpha-amylases of Bacillus stearoth

A;Reference number: A91817; MUID:86195857; PMID:3009417
 A;Accession: B24549

A;Molecule type: DNA
 A;Residues: 1-338, 'G', 340-348, 'S', 350-512 <GRA>

A;Cross-references: GB:IM13256; NID:gl42510; PIDN:AAA22240.1; PID:gl42511
 A;Experimental source: NCIB 8061

R;Stephens, M.A.; Ortlepp, S.A.; Ollington, J.F.; McConnell, D.J.
 J. Bacteriol. 158, 369-372, 1984

A;Title: Nucleotide sequence of the 5' region of the Bacillus licheniformis alpha-amyle

A;Reference number: A91796; MUID:84185455; PMID:6609154

A;Accession: A91796

A;Molecule type: DNA
 A;Residues: 1-104 <STE>

A;Cross-references: GB:K01984; NID:gl42432; PIDN:AAA22193.1; PID:gl42433
 R;Sibakov, M.; Palva, I.

Eur. J. Biochem. 145, 567-572, 1984

A;Title: Isolation and the 5'-end nucleotide sequence of Bacillus licheniformis alpha-a

A;Reference number: A21663; MUID:85076654; PMID:6334606

A;Accession: A21663

A;Molecule type: DNA
 A;Residues: 1-3, 'H', 5-12, 'P', 14-47, 'R', 49-61, 'V', 63, 'D', 65-67, 'VA', 70-71, 'S', 73-80, 'D',

A;Experimental source: chromosomal DNA of ATCC 14580

A;Note: the authors translated the codon CGT for residue 48 as Gly and GAC for residue

R;Laioide, B.M.; Chambliss, G.H.; McConnell, D.J.

J. Bacteriol. 171, 2435-2442, 1989

A;Title: Bacillus licheniformis alpha-amylase gene, amyL, is subject to promoter-indep

A;Reference number: I39773; MUID:89213924; PMID:2540150

A;Accession: I39774

A;Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA
A:Residues: 1-29 <L>AO>
A:Cross-references: GB:M26412; NID:G341477; PIDN:AAA22237.1; PID:G516590
R:Jorgensen, P.L.; Hansen, C.K.; Poulsen, G.B.; Diderichsen, B.
Gene 96, 37-41, 1990
A:Title: In vivo genetic engineering: homologous recombination as a tool for plasmid construction
A:Reference number: 139772; MUID:91092499; PMID:2265757
A:Accession: I39772
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-32, 'I' <J>OR>
A:Cross-references: GB:M62637; NID:G142498; PIDN:AAA22232.1; PID:G142499
R:Kuhn, H.; Fietzek, P.P.; Lampen, J.O.
J. Bacteriol. 149, 372-373, 1982
A:Title: N-terminal amino acid sequence of Bacillus licheniformis alpha-amylase: comparison with the sequence of the precursor
A:Reference number: A26151; MUID:82098050; PMID:6172418
A:Accession: A26151
A:Molecule type: protein
A:Residues: 30-37, 'E', '39-41', 'X', '43-47' <KUH>
R:Machius, M.; Wiegand, G.; Huber, R.
J. Mol. Biol. 246, 545-559, 1995
A:Title: Crystal structure of calcium-depleted Bacillus licheniformis alpha-amylase at 2.2 angstroms, residues 32-210/222-511
A:Reference number: S53788; MUID:95182462; PMID:7877175
A:Accession: S53788
A:Molecule type: protein
A:Residues: 'D', '220-227' <MAC>
A:Note: sequence represents amino end of an internal fragment created by a single enzyme
R:Machius, M.; Wiegand, G.; Huber, R.
submitted to the Brookhaven Protein Data Bank, July 1995
A:Reference number: A52206; PDB:1BPL
A:Contents: annotation; X-ray crystallography, 2.2 angstroms, residues 32-210/222-511
A:Note: these structural studies suggest 163 is leu rather than Arg
R:Song, H.K.; Hwang, K.Y.; Chang, C.; Suh, S.W.
submitted to the Brookhaven Protein Data Bank, October 1996
A:Reference number: A56860; PDB:1VJS
A:Contents: annotation; X-ray crystallography, 1.7 angstroms, residues 32-210/222-511
C:Genetics:
A:Gene: amyL
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Superfamily: alpha-amylase, amyloliquefaciens type; alpha-amylase core homology
C:Keywords: extracellular protein; glycosidase; heat-stable protein; hydrolase; polysaccharide
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-512/Product: alpha-amylase #status experimental <MAT>
F:227-360/Domain: alpha-amylase core homology <AMY>
F:133,229,264/Binding site: calcium (Asn, Asp, His) #status experimental
F:260,290,357/Active site: Asp, Glu, Asp #status experimental

Query Match 65.6%; Score 42; DB 1; Length 512;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
Db 183 KWHWY 187
|||||

RESULT 14
ALBSN
alpha-amylase (EC 3.2.1.1) precursor - Bacillus amyloliquefaciens
N:Alternate names: 1,4-alpha-D-glucan glucanohydrolase
C:Species: Bacillus amyloliquefaciens
C:Date: 30-Nov-1980 #sequence revision 30-Jun-1987 #text_change 18-Jun-1999
C:Accession: A92389; A90307; I39756; I39763; A00843
R:Takkinen, K.; Pettersson, R.F.; Kalkkinen, N.; Palva, I.; Soderlund, H.; Kaariainen, I.
J. Biol. Chem. 258, 1007-1013, 1983
A:Title: Amino acid sequence of alpha-amylase from Bacillus amyloliquefaciens deduced from cDNA
A:Reference number: A92389; MUID:83108808; PMID:6185474
A:Contents: PUB110
A:Accession: A92389
A:Molecule type: DNA
A:Residues: 1-514 <TAK>

A:Cross-references: GB:J01542; GB:J01543; GB:M12033; GB:M12034; NID:G142428; PIDN:AAA2:
R:Chung, H.S.; Friedberg, F.
A:Biochem. J. 185, 387-395, 1980
A:Title: Sequence of the N-terminal half of Bacillus amyloliquefaciens alpha-amylase.
A:Reference number: A90307; MUID:80241725; PMID:6156671
A:Accession: A90307
A:Molecule type: protein
A:Residues: 32-53, 'I', '55-63', 'L', '65-78', 'D', '80-83', 'S', '85-222' <CHU>
R:Palva, I.; Pettersson, R.F.; Kalkkinen, N.; Lehtovaara, P.; Sarvas, M.; Soderlund, H.
Gene 15, 43-51, 1981
A:Title: Nucleotide sequence of the promoter and NH2-terminal signal peptide region of Bacillus amyloliquefaciens alpha-amylase
A:Reference number: I39756; MUID:82051296; PMID:6170539
A:Accession: I39756
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-96 <RES>
A:Cross-references: EMBL:V00092; NID:G39297; PIDN:CAA23430.1; PID:G39298
R:Ruohonen, L.; Hackman, P.; Lehtovaara, P.; Knowles, J.K.C.; Karanen, S.
Gene 59, 161-170, 1987
A:Title: Efficient secretion of Bacillus amyloliquefaciens alpha-amylase cells by its (alpha-amylase) gene
A:Reference number: I39763; MUID:88137952; PMID:2830166
A:Accession: I39763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-39 <RE2>
A:Cross-references: GB:M18424; NID:G142430; PIDN:AAA22192.1; PID:G142431
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Superfamily: alpha-amylase, amyloliquefaciens type; alpha-amylase core homology
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-31/Domain: signal sequence #status predicted <SIG>
F:32-514/Product: alpha-amylase #status predicted <MPT>
F:229-362/Domain: alpha-amylase core homology <AMY>
F:133,231,266/Binding site: calcium (Asn, Asp, His) #status predicted
F:262,292,359/Active site: Asp, Glu, Asp #status predicted

Query Match 65.6%; Score 42; DB 1; Length 514;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
Db 183 KWHWY 187
|||||

RESULT 15
BHTLE
hemocyanin chain e - American tarantula (Eurypelma californica)
C:Species: Eurypelma californica
C:Date: 15-Nov-1984 #sequence revision 31-Mar-1992 #text_change 22-Jun-1999
C:Accession: S06701; A35772; S13441; A02566; A24179; B37975; S08107
R:Voil, W.M.M.
submitted to the EMBL Data Library, August 1989
A:Reference number: S06701
A:Accession: S06701
A:Molecule type: DNA
A:Residues: 1-624 <VOL>
A:Cross-references: EMBL:X16650; NID:G9249; PIDN:CAA34643.1; PID:G829205
R:Voil, W.; Voil, R.
Proc. Natl. Acad. Sci. U.S.A. 87, 5312-5316, 1990
A:Title: Characterization of the gene encoding the hemocyanin subunit e from the tarantula Eurypelma californica
A:Reference number: A35772; MUID:90319102; PMID:2371273
A:Accession: A35772
A:Molecule type: DNA
A:Residues: 1-64:604-624 <VOL2>
R:Voil, R.; Feldmaier-Fuchs, G.
J. Biol. Chem. 265, 19447-19452, 1990
A:Title: Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs encoding the hemocyanin subunit e from the tarantula Eurypelma californica
A:Reference number: A37975; MUID:91060544; PMID:2246235
A:Accession: S13441
A:Molecule type: mRNA
A:Residues: 1-624 <VOL>

A;Cross-references: EMBL:X16894; NID:99268; PIDN:CAA34772.1; PID:99269
R;Schneider, H.J.; Drexel, R.; Feldmaier, G.; Linzen, B.; Lottspeich, F.; Henschen, A.
Hoppe-Seyler's Z. Physiol. Chem. 364, 1357-1381, 1983
A;Title: Hemocyanins in spiders, XVIII[1]. Complete amino-acid sequence of subunit e from
A;Reference number: A02566; MUID:84059635; PMID:6357986
A;Accession: A02566
A;Molecule type: protein
A;Residues: 2-29, R', 31-46, D', 48-79, K', 81-90, H', 92-110, 112-159, K', 161-216, H', 218-259
C;Genetics:
A;Introns: 65/1; 137/3; 206/2; 338/3; 381/3; 426/3; 489/1; 560/3
C;Superfamily: hemocyanin
C;Keywords: copper; hemolymph; oxygen carrier
F;2-624/Product: hemocyanin chain e #status experimental <MAT>
F;2-150/Domain: 1 <D01>
F;151-374/Domain: 2 <D02>
F;375-624/Domain: 3 <D03>
F;169,173,200/Binding site: copper (His) #status predicted
F;320,324,360/Binding site: copper (His) #status predicted

Query Match 65.6%; Score 42; DB 1; Length 624;
Best Local Similarity 57.1%; Pred. NO. 61;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHYV 9
| | | | |
Db 168 HHWHHI 174

Search completed: December 16, 2003, 14:19:17
Job time : 11.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:35 ; Search time 5.66667 Seconds
(without alignments)
74.689 Million cell updates/sec

Title: US-09-870-089B-9

Perfect score: 64

Sequence: 1 FLKWHWYV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	70.3	153	1 MRGA_BACSU	P37960 bacillus su
2	44	68.8	144	1 G20U_BACSU	P80879 bacillus su
3	43	67.2	299	1 BAH_STRHY	Q01109 streptomyc
4	43	67.2	628	1 HCVC_EURCA	Q9nf16 eurytelma c
5	43	67.2	630	1 HCVC_EURCA	P14750 eurytelma c
6	42	65.6	512	1 AMY_EACLI	P06278 bacillus li
7	42	65.6	514	1 AMY_EACAM	P00692 bacillus am
8	42	65.6	623	1 HCYE_EURCA	P02242 eurytelma c
9	42	65.6	626	1 HCYE_ANDAU	P80476 androctonus
10	42	65.6	628	1 HCYF_EURCA	Q9nf15 eurytelma c
11	40	62.5	172	1 VIF_SIVGB	P22383 simian immu
12	40	62.5	374	1 PAR3_HUMAN	O00254 homo sapien
13	40	62.5	626	1 HCYB_EURCA	Q9nfh9 eurytelma c
14	40	62.5	626	1 HCYD_EURCA	P02241 eurytelma c
15	40	62.5	628	1 HCY2_LIMPO	P04253 limulus pol
16	40	62.5	2126	1 PKDR_MOUSE	Q9z0t6 mus musculu
17	39	60.9	529	1 C6T1_DROME	Q9vri9 drosophila
18	39	60.9	544	1 ALG6_YEAST	Q12001 saccharomyc
19	39	60.9	684	1 TC10_YEAST	P50273 saccharomyc
20	39	60.9	965	1 PBD1_CABEL	Q18115 caenorhabdi
21	38	59.4	176	1 TPFI_TREPA	P16665 treponema p
22	38	59.4	176	1 TPFI_TREPA	P16665 treponema p
23	38	59.4	282	1 PCPF_PBEA9	Q52449 pseudanaba
24	38	59.4	339	1 SPFG_BACSU	P39627 bacillus su
25	38	59.4	512	1 FEN2_YEAST	P25621 saccharomyc
26	37	57.8	294	1 DEH1_MORSP	Q01398 moraxella s
27	37	57.8	515	1 MAD3_YEAST	P47074 saccharomyc
28	37	57.8	628	1 HCYF_EURCA	Q9nf14 eurytelma c
29	37	57.8	661	1 ACSA_COCCI	O13440 coprinus ci
30	37	57.8	675	1 ACSA_CANAL	Q94049 candida alb
31	37	57.8	1002	1 VGNM_CPSMV	P31630 cowpea seve
32	36.5	57.0	220	1 NADD_YERPE	Q8z3g1 versinia pe
33	36	56.2	275	1 TRY1_PIG	Q9n2d1 sus scrofa

RESULT 1

MRGA_BACSU
ID MRGA_BACSU STANDARD; PRT; 153 AA.

AC P37960;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Metalloregulation DNA-binding stress protein.

GN MRGA.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI TaxID=1423;

RN [1]

RP SEQUENCE OF 1-68 FROM N.A.

RX MEDLINE=93374837; PubMed=8396117;

Chen L., James L.P., Helmann J.D.;

"Metalloregulation in Bacillus subtilis: isolation and

characterization of two genes differentially repressed by metal

ions.";

RL J. Bacteriol. 175:5428-5437(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=98015415; PubMed=9359331;

Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;

"Sequencing of regions downstream of adda (98 degrees) and citG (289

degrees) in Bacillus subtilis.";

RL Microbiology 143:3305-3308(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

Denizot F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,

Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,

Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,

Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

Kobayashi Y., Kottler P., Koningstein G., Krohn S., Kumano M.,

Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

Lee S.N., Levine A., Liu H., Masuda S., Maue C., Medigue C.,

Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

Rieger M., Rivolta C., Roche E., Roche M., Rose M., Sadale Y.,

Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,

Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,

Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

34 36 56.2 282 1 BACH_HALSD
35 36 56.2 380 1 YCX7_ODOSI
36 36 56.2 511 1 MATK_PHLPR
37 36 56.2 511 1 MATK_POAPR
38 36 56.2 566 1 PAP_SCHFO
39 36 56.2 686 1 PREI_YEAST
40 36 56.2 718 1 CAT3_NEUCR
41 36 56.2 719 1 CAT3_NEUCR
42 36 56.2 720 1 PFTA_PSEAE
43 36 56.2 745 1 FSP1_RAT
44 36 56.2 1018 1 M2B2_MOUSE
45 36 56.2 1043 1 SYI_CHLEN

ALIGNMENTS

093742 halorubrum
P49833 odontella s
Q9mu24 phleum prat
Q9mu29 poa pratens
Q10285 schizosacch
P32791 saccharomyc
Q9c169 erysiphe gr
P42512 pseudomona
P42512 pseudomona
Q61517 rattus norv
O54782 mus musculu
Q92972 chlamydia p

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*";
 RL Nature 390:249-256(1997).
 RN [4]
 RN CHARACTERIZATION.
 RP MEDLINE=96296451; PubMed=8709848;
 RX Chen L., Helmann J.D.;
 RA "Bacillus subtilis MrgA is a Dps(PexB) homologue: evidence for
 RT metal regulation of an oxidative-stress gene";
 RL Mol. Microbiol. 18:295-300(1995).
 CC -1- FUNCTION: FORMS HIGHLY STABLE, MULTIMERIC PROTEIN-DNA COMPLEXES
 CC WHICH ACCUMULATE IN STATIONARY-PHASE CELLS AND PROTECT AGAINST
 CC OXIDATIVE KILLING.
 CC -1- INDUCTION: BY OXIDATIVE STRESS AND BY GROWTH IN MINIMAL MEDIUM
 CC LACKING IRON [FE-II/III], OR ONE OF THE DIVALENT CATIONS MANGANESE,
 CC COPPER OR COBALT.
 CC -1- SIMILARITY: BELONGS TO THE DPS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L19547; AAA68042.1; -;
 DR EMBL; Z22928; CAA80510.1; -;
 DR EMBL; Z33941; CAB07970.1; -;
 DR EMBL; Z99120; CAB15288.1; -;
 DR PIR; G69660; G69660.
 DR Subtilist; BG10864; mrgA.
 DR InterPro; IPR002177; DPS.
 DR InterPro; IPR001519; Ferritin.
 DR Pfam; PF00210; ferritin; 1.
 DR PRINTS; PR01346; HELNAPAPROT.
 DR ProDom; PD149803; DPS; 1.
 DR PROSITE; PS00818; DPS; 1; 1.
 DR PROSITE; PS00819; DPS; 2; 1.
 KW DNA-binding; Complete proteome.
 SQ SEQUENCE 153 AA; 17332 MW; 3082CF803401E17D CRC64;
 Query Match 70.3%; Score 45; DB 1; Length 153;
 Best Local Similarity 75.0%; Pred. No. 3.4;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LHKWHYV 9
 DB 31 LHRFWYV 38
 RESULT 2
 ID G20U_BACSU STANDARD; PRT; 144 AA.
 AC P80879;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE General stress protein 20U (GSP20U) (DPS protein homolog).
 GN DPS.
 OS *Bacillus subtilis*.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98048467; PubMed=9387221;
 RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
 RT "Sequencing and functional annotation of the *Bacillus subtilis* genes
 RT in the 200 kb *rrnB-dnaB* region.";

Microbiology 143:3431-3441(1997).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Hage K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*";
 RL Nature 390:249-256(1997).
 [3]
 RN SEQUENCE OF 1-14.
 RP STRAIN=168 / IS58;
 RX MEDLINE=97443988; PubMed=9298659;
 RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
 RA Hecker M.;
 RT "First steps from a two-dimensional protein index towards a response-
 RT regulation map for *Bacillus subtilis*.";
 RL Electrophoresis 18:1451-1463(1997).
 CC -1- INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE
 CC LIMITATION AND OXYGEN LIMITATION.
 CC -1- SIMILARITY: BELONGS TO THE DPS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF008220; AAC00237.1; -;
 DR EMBL; Z99119; CAB15043.1; -;
 DR PIR; H69618; H69618.
 DR Subtilist; BG12584; dps.
 DR InterPro; IPR002177; DPS.
 DR InterPro; IPR001519; Ferritin.
 DR Pfam; PF00210; ferritin; 1.
 DR PRINTS; PR01346; HELNAPAPROT.
 DR ProDom; PD149803; DPS; 1.
 DR PROSITE; PS00818; DPS; 1; 1.
 DR PROSITE; PS00819; DPS; 2; 1.
 KW Heat shock; Complete proteome.
 FT INIT MET 0
 SQ SEQUENCE 144 AA; 16462 MW; 21AUBC4438E5E3E2 CRC64;
 Query Match 68.8%; Score 44; DB 1; Length 144;

Best Local Similarity 75.0%; Pred. No. 4.5;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHYV 9
Db 22 LKNYHWYV 29

RESULT 3
BAH_STRHY STANDARD; PRT; 299 AA.
AC Q01109;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acetyl-hydrolase (EC 3.1.1.-).
GN BAH.
OS Streptomyces hygroscopicus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1912;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 21705;
RX MEDLINE=91294191; PubMed=2066341;
RA Raibaud A., Zalacain M., Holt T.G., Tizard R., Thompson C.J.;
RT "Nucleotide sequence analysis reveals linked N-acetyl hydrolase,
RT thioesterase, transport, and regulatory genes encoded by the
RT Bialaphos biosynthetic gene cluster of Streptomyces hygroscopicus.";
RL J. Bacteriol. 173:4454-4463(1991).
CC -!- FUNCTION: THIS PROTEIN REMOVES THE N-ACETYL GROUP FROM
CC BIALAPHOS AS ONE OF THE FINAL STEPS OF THE BIALAPHOS
CC BIOSYNTHETIC PATHWAY.
CC -!- PATHWAY: Bialaphos biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE "GDHG" FAMILY OF LIPOLYTIC ENZYMES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M64783; AAA79277.1;
CC InterPro; IPR002168; Lipolytic_enzyme.
CC InterPro; IPR000379; Ser_estr_site.
CC PROSITE; PS01173; LIPASE_GDYG_HIS; 1.
CC PROSITE; PS01174; LIPASE_GDYG_SER; 1.
CC Hydrolase; Antibiotic biosynthesis.
KW ACT_SITE 73 73 POTENTIAL.
FT ACT_SITE 143 143
SQ SEQUENCE 299 AA; 32096 MW; 4265C8E6E10FAE97 CRC64;
Query Match 67.2%; Score 43; DB 1; Length 299;
Best Local Similarity 83.3%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 HKWHYV 8
Db 267 HVWHYV 272

RESULT 4
HCYC_EURCA STANDARD; PRT; 628 AA.
ID C9NPL6;
AC Q9NPL6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemocyanin C chain (HcC).
GN HCC.
OS Euryelpma californica (American tarantula).

OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Aphonopelma.
OX NCBI_TaxID=29932;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=20564303; PubMed=10961996;
RA Voit R., Feldmaier-Fuchs G., Schweikardt T., Decker H., Burmester T.;
RT "Complete sequence of the 24-mer hemocyanin of the tarantula Euryelpma
RT californicum. Structure and intramolecular evolution of the
RT subunits.";
RL J. Biol. Chem. 275:39339-39344(2000).
CC -!- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
CC ARTHROPODS.
CC -!- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN
CC DIFFERENT CHAINS IDENTIFIED.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Hemolymph.
CC -!- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN
CC LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A
CC BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN
CC ADDITION TO BINDING OXYGEN.
CC -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ277489; CAB89495.1;
CC HSSP; P04253; 1L1A.
CC InterPro; IPR000896; Hemocyanin.
CC InterPro; IPR005203; hemocyanin_C.
CC InterPro; IPR005204; hemocyanin_N.
CC InterPro; IPR002227; Tyrosinase.
CC Pfam; PF03722; hemocyanin; 1.
CC Pfam; PF03723; hemocyanin_C; 1.
CC Pfam; PF03722; hemocyanin_N; 1.
CC PROSITE; PS00209; HEMOCYANIN_1; 1.
CC PROSITE; PS00210; HEMOCYANIN_2; 1.
CC PROSITE; PS00498; TYROSINASE_2; 1.
CC Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.
FT INIT_MET 0 0 BY SIMILARITY.
FT DISULFID 536 584 BY SIMILARITY.
FT METAL 174 174 COPPER 1 (BY SIMILARITY).
FT METAL 178 178 COPPER 1 (BY SIMILARITY).
FT METAL 205 205 COPPER 1 (BY SIMILARITY).
FT METAL 325 325 COPPER 2 (BY SIMILARITY).
FT METAL 329 329 COPPER 2 (BY SIMILARITY).
FT METAL 365 365 COPPER 2 (BY SIMILARITY).
FT CARBOHYD 450 450 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 628 AA; 72435 MW; EC2F8ED04935DFBD CRC64;
Query Match 67.2%; Score 43; DB 1; Length 628;
Best Local Similarity 71.4%; Pred. No. 24;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 HKWHYV 9
Db 173 HHWHYV 179

RESULT 5
HCYA_EURCA STANDARD; PRT; 630 AA.
ID HCYA_EURCA
AC P14750;
DT 01-APR-1990 (Rel. 14, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemocyanin A chain (HcA).
GN HcA.
OS Eurytelma californica (American tarantula).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Aphonopelma.
OX NCBI_TaxID=29932;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=91060544; PubMed=2246235;
RA Voit R., Feldmaier-Fuchs G.;
RT "Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs
RT encoding the tarantula hemocyanin subunits a and e.";
RL J. Biol. Chem. 265:19447-19452(1990).
RN [2]
RN SEQUENCE.
RP MEDLINE=91025623; PubMed=2222854;
RA Schartau W., Metzger W., Sonner P., Geisert H., Storz H.;
RT "Hemocyanins in spiders, XXIII. Complete amino-acid sequence of
RT subunit a of Eurytelma californicum hemocyanin.";
RL Biol. Chem. Hoppe-Seyler 371:557-565(1990).
CC -!- FUNCTION: HEMOCYANIN IS COPPER-CONTAINING OXYGEN CARRIERS
CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
CC ARTERIOPODS.
CC -!- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN
CC DIFFERENT CHAINS IDENTIFIED.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Hemolymph.
CC -!- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN
CC LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A
CC BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN
CC ADDITION TO BINDING OXYGEN.
CC -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X16893; CAA34771.1; -;
DR HSP; P04253; 1LLA.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN 1; 1.
DR PROSITE; PS00210; HEMOCYANIN 2; 1.
DR PROSITE; PS00498; TYROSINASE 2; 1.
KW Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.
FT INIT MET 0
FT METAL 174 174 COPPER 1 (PROBABLE).
FT METAL 178 178 COPPER 1 (PROBABLE).
FT METAL 205 205 COPPER 1 (PROBABLE).
FT METAL 325 325 COPPER 2 (PROBABLE).
FT METAL 329 329 COPPER 2 (PROBABLE).
FT METAL 365 365 COPPER 2 (PROBABLE).
FT METAL 537 585 BY SIMILARITY.
FT DISULFID 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 630 AA; 72187 MW; 4f69899631CC793 CRC64;
Query Match 67.2%; Score 43; DB 1; Length 630;

Best Local Similarity 71.4%; Pred. No. 24;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 HKWHWYV 9
DB 173 HKWHWYV 179
| | | | |
| | | | |
RESULT 6
ID AMY_BACLI STANDARD; PRT; 512 AA.
AC P06278;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-amylase precursor (SC 3.2.1.1) (1,4-alpha-D-glucan
DE glucanhydrolase).
GN AMYS OR AMYL.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 27811;
RX MEDLINE=86111694; PubMed=2418011;
RA Yuuki T., Nomura T., Tezuka H., Tsuboi A., Yamagata H.,
RA Tsukagoshi N., Uda S.;
RT "Complete nucleotide sequence of a gene coding for heat- and
RT pH-stable alpha-amylase of Bacillus licheniformis: comparison of the
RT amino acid sequences of three bacterial liquefying alpha-amylases
RT deduced from the DNA sequences.";
RL J. Biochem. 98:1147-1156(1985).
RN [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=86195857; PubMed=3009417;
RX Gray G.L., Mainzer S.E., Rey M.W., Lamsa M.H., Kindle K.L.,
RA Carmona C., Requaet C.;
RT "Structural genes encoding the thermophilic alpha-amylases of
RT Bacillus stearothermophilus and Bacillus licheniformis.";
RL J. Bacteriol. 166:635-643(1986).
RN [3]
RP SEQUENCE OF 1-104 FROM N.A.
RX MEDLINE=84185455; PubMed=6609154;
RA Stephens M.A., Ortlepp S.A., Ollington J.F., McConnell D.J.;
RT "Nucleotide sequence of the 5' region of the Bacillus licheniformis
RT alpha-amylase gene: comparison with the B. amyloliquefaciens gene.";
RL J. Bacteriol. 158:369-372(1984).
RN [4]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=89213924; PubMed=2540150;
RA Laoid B.M., Chambliss G.H., McConnell D.J.;
RT "Bacillus licheniformis alpha-amylase gene, amyL, is subject to
RT promoter-independent catabolite repression in Bacillus subtilis.";
RL J. Bacteriol. 171:2435-2442(1989).
RN [5]
RP SEQUENCE OF 30-47.
RX MEDLINE=82098050; PubMed=6172418;
RA Kuhn H., Fietzek P.P., Lampen J.O.;
RT "N-terminal amino acid sequence of Bacillus licheniformis
RT alpha-amylase: comparison with Bacillus amyloliquefaciens and
RT Bacillus subtilis Enzymes.";
RL J. Bacteriol. 149:372-373(1982).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC STRAIN=ATCC 27811;
RX MEDLINE=95182462; PubMed=7877175;
RA Machius M., Wiegand G., Huber R.;
RT "Crystal structure of calcium-depleted Bacillus licheniformis alpha-
RT amylase at 2.2-A resolution.";
RL J. Mol. Biol. 246:545-559(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -!- BIOTECHNOLOGY: Used in the food industry for high temperature

liquefaction of starch-containing mashes and in the detergent industry to remove starch. Sold under the name Termamyl by Novozymes.

-!- MISCELLANEOUS: ABLE TO WORK AT RELATIVELY HIGH (ALKALINE) PH VALUES (UP TO PH 11) AND AT HIGH TEMPERATURES (UP TO 100 DEGREE CELSIUS).

-!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X03236; CAA26981.1; -
 EMBL; M38570; AAA22226.1; -
 EMBL; M13256; AAA22240.1; -
 EMBL; K01984; AAA22193.1; -
 EMBL; M26412; AAA22237.1; -
 EMBL; A17930; CAA01355.1; -
 PIR; A91997; ALBSL
 PDB; 1BPL; 17-AUG-96.
 PDB; 1VJS; 12-MAR-97.
 PDB; 1BLI; 23-MAR-99.
 PDB; 1E3X; 21-JUN-01.
 PDB; 1E3Z; 21-JUN-01.
 PDB; 1E40; 21-JUN-01.
 PDB; 1E43; 21-JUN-01.
 PDB; 1O80; 30-JAN-03.
 InterPro; IPR006589; Alp_ amyl_cat_sub.
 InterPro; IPR006047; Alpha_ amyl_cat.
 InterPro; IPR006046; Glyco_hydro_13.
 Pfam; PF00128; alpha-amylase; 1.
 PRINTS; PR00110; ALPHAAMYLASE.
 SMART; SMO0642; Amy; 1.
 Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; 3D-structure.

1 29
 SIGNAL
 CHAIN 30 512
 ACT SITE 260 260
 ACT SITE 264 264
 ACT SITE 357 357
 ACT SITE 357 357
 CONFLICT 38 163
 CONFLICT 339 339
 CONFLICT 349 349
 STRAND 36 38
 TURN 42 43
 TURN 42 43
 TURN 50 56
 TURN 57 57
 TURN 58 63
 TURN 64 65
 STRAND 68 70
 STRAND 76 76
 TURN 80 81
 STRAND 89 89
 STRAND 90 91
 STRAND 98 98
 TURN 99 100
 STRAND 101 101
 STRAND 109 109
 TURN 122 123
 TURN 122 123
 STRAND 125 130
 STRAND 134 135
 STRAND 140 149
 TURN 154 155
 STRAND 157 157
 STRAND 162 170
 TURN 173 177
 STRAND 184 184
 HELIX 186 188

189 192
 195 195
 199 199
 200 200
 204 206
 223 226
 228 230
 232 233
 233 233
 235 252
 253 253
 253 253
 256 259
 264 264
 267 281
 282 282
 286 289
 296 305
 306 308
 311 313
 315 326
 327 329
 333 337
 341 344
 346 348
 349 350
 352 352
 356 358
 360 361
 363 364
 370 372
 370 372
 373 382
 389 391
 392 396
 410 422
 423 423
 431 431
 437 442
 445 445
 446 447
 448 448
 449 450
 453 458
 463 468
 471 473
 474 475
 477 480
 481 482
 489 490
 493 494
 496 501
 503 504
 506 510
 512 AA; 58549 MW; D8BB77759CD4C482 CRC64;
 SQ SEQUENCE

Query Match 65.6%; Score 42; DB 1; Length 512;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KWHWY 8
 Db 183 KWHWY 187

RESULT 7
 AMY_BACAM
 ID AMY_BACAM STANDARD; PRT; 514 AA.
 AC P00692;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alpha-amylase precursor (EC 3.2.1.1) (1,4-alpha-D-glucan
 DE glucanohydrolase).
 OS Bacillus amyloliquefaciens.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

Matches	5;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps
QY	4	KWHWY 8 						
Db	183	KWHWY 187						
RESULT 8								
ID	HCYE_EURCA	STANDARD;	PRT;	623 AA.				
AC	F02242;							
DT	21-JUL-1986	(Rel. 01, Created)						
DT	01-APR-1990	(Rel. 14, Last sequence update)						
DT	28-FEB-2003	(Rel. 41, Last annotation update)						
DE	Hemocyanin E chain [HcE].							
GN	HCE.							
OS	Eurypelma californica (American tarantula).							
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata;							
OC	Cygalomorphae; Theraphosidae; Aphonopelma.							
NB	NCBI_TaxID=29932;							
RN	[1]_TaxID=29932;							
RP	SEQUENCE FROM N.A.							
RA	MEDLINE=90319102; PubMed=2371273;							
RX	Voll W., Voit R.;							
RT	"Characterization of the gene encoding the hemocyanin subunit e from							
RL	the tarantula Eurypelma californicum."							
RN	Proc. Natl. Acad. Sci. U.S.A. 87:5312-5316(1990).							
RP	SEQUENCE.							
RA	MEDLINE=84059635; PubMed=6357986;							
RX	Schneider H.-J.; Drexler R.; Feldmaier G.; Linzen B.; Lottspeich F.;							
RT	Henschen A.;							
RL	"Hemoglobins in Spiders, XVIII. Complete amino-acid sequence of							
RN	hemoglobin epsilon from Eurypelma californicum hemocyanin.";							
RP	Hoppe-Seyler's Z. Physiol. Chem. 364:1357-1381(1983).							
RA	SEQUENCE FROM N.A.							
RX	MEDLINE=91060544; PubMed=2246235;							
RT	Voit R.; Feldmaier-Fuchs G.;							
RL	"Arthropod hemoglobins. Molecular cloning and sequencing of cDNAs							
RN	encoding the tarantula hemocyanin subunits alpha and epsilon";							
RP	J. Biol. Chem. 265:19447-19452(1990).							
RA	SEQUENCE OF 74-599 FROM N.A.							
RX	MEDLINE=86300721; PubMed=3017715;							
RT	Voit R.; Schneider H.-J.;							
RL	"Tarantula hemocyanin mRNA. In vitro translation, cDNA cloning and							
RN	nucleotide sequence corresponding to subunit epsilon";							
RP	Eur. J. Biochem. 159:23-29(1986).							
CC	-1- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS							
CC	OCCURRING FREELY DISSOLVED IN THE HEMOLYPH OF MANY MOLLUSKS AND							
CC	ARTHROPODS.							
CC	-1- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN							
CC	DIFFERENT CHAINS IDENTIFIED.							
CC	-1- SUBCELLULAR LOCATION: Extracellular.							
CC	-1- TISSUE SPECIFICITY: Hemolymph.							
CC	-1- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN							
CC	LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A							
CC	BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN							
CC	ADDITION TO BINDING OXYGEN.							
CC	-1- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN							
CC	SUBFAMILY.							
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration							
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation							
CC	at the European Bioinformatics Institute. There are no restrictions on its							
CC	use by non-profit institutions as long as its content is in no way							
CC	modified and this statement is not removed. Usage by and for commercial							
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/							
CC	or send an email to licens@isb-sib.ch).							
DR	ENBL; X16894; CAA34772.1; -.							
DR	ENBL; X04291; CAA37838.1; -.							

DR EMBL; X16650; CAA34643.1; JOINED.
DR EMBL; X16651; CAA34643.1; JOINED.
DR EMBL; X16652; CAA34643.1; JOINED.
DR EMBL; X16653; CAA34643.1; JOINED.
DR EMBL; X16654; CAA34643.1; JOINED.
DR EMBL; X16655; CAA34643.1; JOINED.
DR EMBL; X16656; CAA34643.1; JOINED.
DR EMBL; X16657; CAA34643.1; JOINED.
DR PIR; S06701; BHTLE.
DR HSP; P04253; IOXY.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
DR Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.
FT INIT_MET 0
FT METAL 168 168 COPPER 1 (PROBABLE).
FT METAL 172 172 COPPER 1 (PROBABLE).
FT METAL 199 199 COPPER 1 (PROBABLE).
FT METAL 319 319 COPPER 2 (PROBABLE).
FT METAL 323 323 COPPER 2 (PROBABLE).
FT METAL 359 359 COPPER 2 (PROBABLE).
FT METAL 528 528 BY SIMILARITY.
FT DISULFID 528 576 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 444 444
FT CONFLICT 29 29 D -> R (IN REF. 2).
FT CONFLICT 46 46 C -> D (IN REF. 2).
FT CONFLICT 79 79 R -> K (IN REF. 2).
FT CONFLICT 90 90 A -> H (IN REF. 2).
FT CONFLICT 110 110 MISSING (IN REF. 2).
FT CONFLICT 159 159 K -> R (IN REF. 1 AND 3).
FT CONFLICT 216 216 R -> H (IN REF. 2).
FT CONFLICT 254 254 H -> M (IN REF. 2).
FT CONFLICT 307 307 F -> H (IN REF. 2).
FT CONFLICT 325 325 M -> MK (IN REF. 2).
FT CONFLICT 528 530 CGW -> DGK (IN REF. 2).
FT CONFLICT 561 563 NGH -> D (IN REF. 2).
FT CONFLICT 602 602 F -> L (IN REF. 3).
FT CONFLICT 608 608 V -> P (IN REF. 3).
SQ SEQUENCE 623 AA; 71545 MW; E411052A80814004 CRC64;

Query Match 65.6%; Score 42; DB 1; Length 623;
Best Local Similarity 57.1%; Pred.No. 33;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 HKWHYV 9
Db 167 HHWHWHI 173

RESULT 9
HCYF ANDAU
ID HCYF ANDAU STANDARD; PRT; 626 AA.
AC P04076;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemocyanin AA6 chain.
OS Androctonus australis hector (Sahara scorpion).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC Butchoidea; Butidae; Androctonus.
OX NCBI_TaxID=70175;
RN [1]
RP SEQUENCE.
RC TISSUE=Hemolymph;
EX MEDLINE=96061936; PubMed=7588779;
RA Buzi A., Gagnon J., Lamy J., Thibault P., Forest E.,
Hudry-Clergeon G.;

RT "Complete amino acid sequence of the Aa6 subunit of the scorpion
RT Androctonus australis hemocyanin determined by Edman degradation and
RT mass spectrometry";
RL Eur. J. Biochem. 233:93-101(1995).
CC -!- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS AND
CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
CC ARTHROPODS.
CC -!- SUBUNIT: SCORPION HEMOCYANIN IS A 24-CHAIN POLYMER WITH 8
CC DIFFERENT CHAINS IDENTIFIED, ASSEMBLED IN HEXAMERIC SUBSTRUCTURES.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Hemolymph.
CC -!- PTM: THREE DISULFIDE BONDS ARE PRESENT.
CC -!- MASS SPECTROMETRY: MW=71890; MW ERR=7; METHOD=Electrospray.
CC -!- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN
CC LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A
CC BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN
CC ADDITION TO BINDING OXYGEN.
CC -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
CC SUBFAMILY.

DR PIR; S67964; S67964.
DR HSP; P04253; IOXY.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
DR Transport; Oxygen transport; Copper; Hemolymph;
KW Phosphorylation.
FT METAL 170 170 COPPER 1 (PROBABLE).
FT METAL 174 174 COPPER 1 (PROBABLE).
FT METAL 201 201 COPPER 1 (PROBABLE).
FT METAL 321 321 COPPER 2 (PROBABLE).
FT METAL 325 325 COPPER 2 (PROBABLE).
FT METAL 361 361 COPPER 2 (PROBABLE).
FT MOD_RES 374 374 PHOSPHORYLATION (PROBABLE).
SQ SEQUENCE 626 AA; 71785 MW; E788136AE3DEF0D2 CRC64;

Query Match 65.6%; Score 42; DB 1; Length 626;
Best Local Similarity 57.1%; Pred.No. 33;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 HKWHYV 9
Db 169 HHWHWHI 175

RESULT 10
HCYF EURCA
ID HCYF EURCA STANDARD; PRT; 628 AA.
AC Q9NFI5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemocyanin F chain (Hcf).
GN HCF.
OS Eurytelma californica (American tarantula).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Aphonopelma.
OX NCBI_TaxID=29932;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=20564303; PubMed=10961996;
RA Voit R., Feldmaier-Fuchs G., Schweikardt T., Decker H., Burnester T.;
RT "Complete sequence of the 24-mer hemocyanin of the tarantula Eurytelma
RT californicum. Structure and intramolecular evolution of the
RT subunits.";

```

RL J. Biol. Chem. 275:39339-39344 (2000).
CC -!- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
CC ARTHROPODS.
CC
CC -!- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN
CC DIFFERENT CHAINS IDENTIFIED.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Hemolymph.
CC -!- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN
CC LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A
CC BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN
CC ADDITION TO BINDING OXYGEN.
CC -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ277491; CAB89496.1; -.
CC HSSP; P04253; 1LLA.
CC InterPro; IPR000896; Hemocyanin.
CC InterPro; IPR005203; hemocyanin_C.
CC InterPro; IPR005204; hemocyanin_N.
CC InterPro; IPR002227; Tyrosinase.
CC Pfam; PF00372; hemocyanin; 1.
CC Pfam; PF03722; hemocyanin_C; 1.
CC Pfam; PF03723; hemocyanin_N; 1.
CC PROSITE; PS00209; HEMOCYANIN 1; 1.
CC PROSITE; PS00210; HEMOCYANIN 2; 1.
CC PROSITE; PS00498; TYROSINASE 2; 1.
CC TransPort; OX098; Copper; Glycoprotein; Hemolymph.
CC INIT MET 0
CC BY SIMILARITY.
FT DISULFID 533 581
FT METAL 171 171
FT METAL 175 175
FT METAL 202 202
FT METAL 323 323
FT METAL 327 327
FT METAL 363 363
FT METAL 394 394
FT CARBOHYD 446 446
FT CARBOHYD 526 526
FT CARBOHYD 614 614
FT SEQUENCE 628 AA; 72017 MW; 539C94849CC8D4F4 CRC64;
SQ
Query Match 65.6%; Score 42; DB 1; Length 628;
Best Local Similarity 57.1%; Pred. No. 33;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKHWHYV 9
DB 170 HHWHHI 176

RESULT 11
VIF_SIVGB
ID VIF_SIVGB STANDARD; PRT; 172 AA.
AC P22383;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Virion infectivity factor (SOR protein) (Q protein).
GN VIF.
OS Simian immunodeficiency virus (isolate GB1).
OC Viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11732;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=90015168; PubMed=2797181;
RA Tsujimoto H., Hasegawa A., Maki N., Fukasawa M., Miura T., Speidel S.,
RA Cooper K.W., Moriyama E.N., Gojobori T., Hayami M.;
RT "Sequence of a novel simian immunodeficiency virus from a wild-caught
RT African mandrill.";
RL Nature 341:539-541 (1989).
CC -!- FUNCTION: DETERMINES VIRUS INFECTIVITY.
CC -!- MISCELLANEOUS: THIS IS AN AFRICAN MANDRILL ISOLATE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M27470; AAB49570.1; -.
CC InterPro; IPR000475; Viral_infect.
CC Pfam; PF00559; Vif; 1.
CC PRINTS; PR00349; VIRIONINFECT.
CC ProDom; PD000063; Viral_infect; 1.
CC AIDS.
CC SEQUENCE 172 AA; 20499 MW; 205E7BDD6265FCB CRC64;
SQ
Query Match 62.5%; Score 40; DB 1; Length 172;
Best Local Similarity 62.5%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LHKHWHYV 9
DB 19 LKHEWLV 26

RESULT 12
PAR3_HUMAN
ID PAR3_HUMAN STANDARD; PRT; 374 AA.
AC O00254;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Proteinase activated receptor 3 precursor (PAR-3) (Thrombin receptor-
DE like 2) (Coagulation factor II receptor-like 2).
GN F2RL2 OR PAR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF THR-39 AND PHE-40.
RX MEDLINE=97242411; PubMed=9087410;
RA Ishihara H., Connolly A.J., Zeng D., Kahn M.L., Zheng Y.-W.,
RA Timmons C., Tram T., Coughlin S.R.;
RT "Protease-activated receptor 3 is a second thrombin receptor in
RT humans.";
RL Nature 386:502-506 (1997).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS SER-15; VAL-177 AND ASP-250.
RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,
RA Nickerson D.A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=98279023; PubMed=9614115;
RA Schmidt V.A., Nierman W.C., Maglott D.R., Cupit L.D., Moskowitz K.A.,
RA Wainer J.A., Bahou W.F.;
RT "The human proteinase-activated receptor-3 (PAR-3) gene.
RT Identification within a PAR gene cluster and characterization in
RT vascular endothelial cells and platelets.";
RL J. Biol. Chem. 273:15061-15068 (1998).
RN [4]
RP FUNCTION.
RP MEDLINE=99178892; PubMed=10079109;

```

RA Kahn M.L., Nakanishi-Matsui M., Shapiro M.J., Ishihara H.,
 RA Coughlin S.R.;
 RT "Protease-activated receptors 1 and 4 mediate activation of human
 RT platelets by thrombin.";
 RL J. Clin. Invest. 103:879-887(1999).
 CC -!- FUNCTION: Receptor for activated thrombin coupled to G proteins
 CC that stimulate phosphoinositide hydrolysis.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Highest expression in the megakaryocytes of
 CC the bone marrow, lower in mature megakaryocytes, in platelets and
 CC in a variety of other tissues such as heart and gut.
 CC -!- PTM: A proteolytic cleavage generates a new amino terminus that
 CC functions as a tethered ligand.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U92971; AAC51218.1; -.
 DR EMBL; AF374726; AAK51564.1; -.
 DR HSSP; P34996; 1DDO.
 DR Genew; HGNC:3539; F2RL2.
 DR MIM; 601919; -.
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0005057; P:thrombin receptor activity; TAS.
 DR GO; GO:0007596; P:blood coagulation; TAS.
 DR GO; GO:0007203; P:phosphatidylinositol-4,5-bisphosphate hydro. .; TAS.
 DR GO; GO:0009611; P:response to wounding; TAS.
 DR InterPro; IPR000276; GPCR_Rhodopsin.
 DR Pfam; PF00001; 7tm 1; 1.
 DR PRINTS; PRO0237; GPCR_Rhodopsin.
 DR PROSITE; PS00237; G-PROTEIN RECP F1_1; 1.
 DR PROSITE; PS00262; G-PROTEIN RECP F1_2; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Blood coagulation; Polymorphism.
 FT SIGNAL 1 21
 FT PROPEP 22 38
 FT CHAIN 39 374
 FT DOMAIN 39 94
 FT TRANSMEM 95 120
 FT DOMAIN 121 128
 FT TRANSMEM 129 148
 FT DOMAIN 149 167
 FT TRANSMEM 168 189
 FT DOMAIN 190 206
 FT TRANSMEM 207 230
 FT DOMAIN 231 260
 FT TRANSMEM 261 280
 FT DOMAIN 281 297
 FT TRANSMEM 298 322
 FT DOMAIN 323 336
 FT TRANSMEM 337 361
 FT DOMAIN 362 374
 FT SITE 38 39
 FT DISULFID 166 245
 FT CARBOHYD 25 25
 FT CARBOHYD 82 82
 FT CARBOHYD 331 331
 FT VARIANT 15 15
 FT VARIANT 177 177
 FT VARIANT 250 250
 FT MUTAGEN 39 39
 FT MUTAGEN 40 40

FT SQ SEQUENCE 374 AA; 42508 MW; C45C15A695DD1ABB CRC64;
 CLEAVAGE.
 Query Match 62.5%; Score 40; DB 1; Length 374;
 Best Local Similarity 71.4%; Pred. No. 40;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 3 HKWHWYV 9
 Db 294 HRWLWIV 300
 RESULT 13
 HCVB EURCA STANDARD; PRT; 626 AA.
 AC Q9NEH9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemocyanin B chain (Hcb).
 GN HCB.
 OS Eurytelma californica (American tarantula).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Mygalomorphae; Theraphosidae; Aphonopelma.
 OX NCBI_TaxID=29932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=20564303; PubMed=10961996;
 RA Voit R., Feldmaier-Fuchs G., Schweikardt T., Decker H., Burmester T.;
 RT "Complete sequence of the 24-mer hemocyanin of the tarantula Eurytelma
 RT californicum. Structure and intramolecular evolution of the
 RT subunits.";
 RL J. Biol. Chem. 275:39339-39344 (2000).
 CC -!- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
 CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
 CC ARTHROPODS.
 CC -!- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN
 CC DIFFERENT CHAINS IDENTIFIED.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: Hemolymph.
 CC -!- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN
 CC LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A
 CC BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN
 CC ADDITION TO BINDING OXYGEN.
 CC -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ290429; CAB89498.1; -.
 DR HSSP; P04253; 1LLA.
 DR InterPro; IPR000896; Hemocyanin.
 DR InterPro; IPR005203; hemocyanin_C.
 DR InterPro; IPR005204; hemocyanin_N.
 DR InterPro; IPR002227; Tyrosinase.
 DR Pfam; PF03723; hemocyanin; 1.
 DR Pfam; PF03723; hemocyanin; 1.
 DR Pfam; PF03722; hemocyanin_N; 1.
 DR PRINTS; PR00187; HAEMOCYANIN.
 DR PROSITE; PS00209; HEMOCYANIN_1; 1.
 DR PROSITE; PS00210; HEMOCYANIN_2; 1.
 DR PROSITE; PS00498; TYROSINASE_2; 1.
 KW Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.
 FT INIT MET 0 0 BY SIMILARITY.
 FT DISULFID 533 581 BY SIMILARITY.
 FT METAL 172 172 COPPER 1 (BY SIMILARITY).

```

FT METAL 176 176 COPPER 1 (BY SIMILARITY).
FT METAL 203 203 COPPER 1 (BY SIMILARITY).
FT METAL 323 323 COPPER 2 (BY SIMILARITY).
FT METAL 327 327 COPPER 2 (BY SIMILARITY).
FT METAL 363 363 COPPER 2 (BY SIMILARITY).
FT METAL 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 626 AA; 71966 MW, 1764FB28C91F9974 CRC64;
SQ SEQUENCE 626 AA; 71966 MW, 1764FB28C91F9974 CRC64;

Query Match 62.5%; Score 40; DB 1; Length 626;
Best Local Similarity 57.1%; Pred. No. 64;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWV 9
DB 171 HHWHHL 177

RESULT 14
HCYD EURCA STANDARD; PRT; 626 AA.
ID P02241; Q9NPH8;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemocyanin D chain (HcD).
GN HCD.
OS Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Aphonopelma.
OX NCBI_taxID=29932;
RN [1]
RP SEQUENCE.
RX MEDLINE=84059636; PubMed=6642428;
RA Schartau W., Everle F., Reisinger P., Geisert H., Storz H., Linzen B.;
RT "Hemocyanins in spiders, XIX. Complete amino-acid sequence of subunit
d from Eurypelma californicum hemocyanin, and comparison to chain e.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:1383-1409(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=20564303; PubMed=10961996;
RA Voit R., Feldmaier-Fuchs G., Schweikardt T., Decker H., Burmester T.;
RT "Complete sequence of the 24-mer hemocyanin of the tarantula Eurypelma
californicum. Structure and intramolecular evolution of the
subunits.";
RL J. Biol. Chem. 275:39339-39344(2000).
CC -1- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
ARTHROPODS.
CC -1- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN
DIFFERENT CHAINS IDENTIFIED.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Hemolymph.
CC -1- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN
LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A
BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN
ADDITION TO BINDING OXYGEN.
CC -1- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ290430; CAB89499.1; ..
CC HSSP; P04253; 1LLA.
CC InterPro; IPR000896; Hemocyanin.
CC InterPro; IPR005203; hemocyanin_C.

```

```

DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin_1.
DR Pfam; PF03722; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
KW Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.
FT INIT MET 0
FT METAL 170 170 COPPER 1 (BY SIMILARITY).
FT METAL 174 174 COPPER 1 (BY SIMILARITY).
FT METAL 201 201 COPPER 1 (BY SIMILARITY).
FT METAL 321 321 COPPER 2 (BY SIMILARITY).
FT METAL 325 325 COPPER 2 (BY SIMILARITY).
FT METAL 361 361 COPPER 2 (BY SIMILARITY).
FT METAL 530 530 BY SIMILARITY.
FT DISULFID 578 N-LINKED (GLCNAC. . .) (Probable).
FT CARBOHYD 444 444 SC -> GS (IN REF. 2).
FT CONFLICT 46 47 ETLAEALVE -> IVFIEIHD (IN REF. 1).
FT CONFLICT 57 65 I -> L (IN REF. 1).
FT CONFLICT 73 73 I -> L (IN REF. 1).
FT CONFLICT 96 96 L -> I (IN REF. 1).
FT CONFLICT 107 107 L -> I (IN REF. 1).
FT CONFLICT 133 133 V -> I (IN REF. 1).
FT CONFLICT 185 185 I -> F (IN REF. 1).
FT CONFLICT 221 221 I -> L (IN REF. 1).
FT CONFLICT 226 226 I -> L (IN REF. 1).
FT CONFLICT 238 238 R -> S (IN REF. 1).
FT CONFLICT 265 265 MISSING (IN REF. 1).
FT CONFLICT 292 292 L -> I (IN REF. 1).
FT CONFLICT 298 301 LNIL -> INVI (IN REF. 1).
FT CONFLICT 316 316 F -> Y (IN REF. 1).
FT CONFLICT 402 403 LI -> IL (IN REF. 1).
FT CONFLICT 418 420 INL -> LNI (IN REF. 1).
FT CONFLICT 428 428 Y -> YNY (IN REF. 1).
FT CONFLICT 470 470 I -> L (IN REF. 1).
FT CONFLICT 530 532 CGW -> DGK (IN REF. 1).
FT CONFLICT 561 561 K -> S (IN REF. 1).
FT CONFLICT 571 572 CS -> DQ (IN REF. 1).
SQ SEQUENCE 626 AA; 71991 MW; P45C4DEB8CE09E2C CRC64;

Query Match 62.5%; Score 40; DB 1; Length 626;
Best Local Similarity 57.1%; Pred. No. 64;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWV 9
DB 169 HHWHHL 175

RESULT 15
HCY2 LIMPO STANDARD; PRT; 628 AA.
ID HCY2 LIMPO
AC P04253;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemocyanin II.
OS Limulus polyphemus (Atlantic horseshoe crab).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC Limulidae; Limulus.
OX NCBI_taxID=6850;
RN [1]
RP SEQUENCE.
RX MEDLINE=86278119; PubMed=3525550;
RA Nakashima H., Böhrens P.Q., Moore M.D., Yokota E., Riggs A.F.;
RT "Structure of hemocyanin II from the horseshoe crab, Limulus
polyphemus. Sequences of the overlapping peptides, ordering the CNBr
fragments, and the complete amino acid sequence.";
RL J. Biol. Chem. 261:10526-10533(1986).
RN [2]
RP SEQUENCE OF 1-203.

```

RX MEDLINE=84185567; PubMed=6715319;
 RA Yokota E., Riggs A.F.;
 RT "The structure of the hemocyanin from the horseshoe crab, Limulus
 polyphemus. The amino acid sequence of the largest cyanogen bromide
 fragment.";
 RL J. Biol. Chem. 259:4739-4749(1984).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY.
 RA MEDLINE=91326804; PubMed=1866430;
 RX Magnus K.A., Lattman E.E., Volbeda A., Hol W.G.J.;
 RT "Hexamers of subunit II from Limulus hemocyanin (a 48-mer) have the
 same quaternary structure as whole Fanulirus hemocyanin molecules.";
 RL Proteins 9:240-247(1991).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.18 ANGSTROMS).
 RA MEDLINE=93299372; PubMed=8518732;
 RX Hazes B., Magnus K.A., Bonaventura C., Bonaventura J., Dauter Z.,
 Kalk K.H., Hol W.G.J.;
 RT "Crystal structure of deoxygenated limulus polyphemus subunit II
 hemocyanin at 2.18-A resolution: clues for a mechanism for allosteric
 regulation.";
 RL Protein Sci. 2:597-619(1993).
 CC -!- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS
 CC -!- OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
 CC ARTHROPODS.
 CC -!- SUBUNIT: HEXAMER OR A MULTIPLE THEREOF.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: Hemolymph.
 CC -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
 CC SUBFAMILY.
 DR PIR; A26713; BHC2A.
 DR PDB; 1LLA; 31-JAN-94.
 DR PDB; 1LL1; 20-AUG-97.
 DR PDB; INOL; 08-MAR-96.
 DR PDB; 1OXY; 27-FEB-95.
 DR InterPro; IPR000896; Hemocyanin.
 DR InterPro; IPR005203; hemocyanin_C.
 DR InterPro; IPR005204; hemocyanin_N.
 DR InterPro; IPR002227; Tyrosinase.
 DR Pfam; PF03723; hemocyanin_C; 1.
 DR Pfam; PF03722; hemocyanin_N; 1.
 DR PRINTS; PR00187; HAEMOCYANIN.
 DR PROSITE; PS00209; HEMOCYANIN_1; 1.
 DR PROSITE; PS00210; HEMOCYANIN_2; 1.
 DR PROSITE; PS00498; TYROSINASE_2; 1.
 KW Oxygen transport; Transport; Copper; Glycoprotein;
 KW Hemolymph; 3D-structure.
 FT MOD_RES 1 1 BLOCKED (PARTIAL).
 FT METAL 173 173 COPPER 1.
 FT METAL 177 177 COPPER 1.
 FT METAL 204 204 COPPER 1.
 FT METAL 324 324 COPPER 2.
 FT METAL 328 328 COPPER 2.
 FT METAL 364 364 COPPER 2.
 FT HELIX 3 13
 FT TURN 14 15
 FT TURN 16 18
 FT HELIX 33 36
 FT TURN 42 43
 FT TURN 51 65
 FT TURN 66 66
 FT TURN 70 80
 FT TURN 81 83
 FT HELIX 86 99
 FT HELIX 101 103
 FT TURN 104 105
 FT HELIX 111 114
 FT HELIX 116 118
 FT TURN 119 119
 FT TURN 122 130
 FT HELIX 156 160
 FT HELIX 161 164

FT TURN 165 165
 FT HELIX 167 179
 FT TURN 182 183
 FT HELIX 186 189
 FT TURN 194 195
 FT HELIX 196 217
 FT TURN 218 219
 FT TURN 229 230
 FT TURN 233 233
 FT STRAND 236 236
 FT STRAND 241 242
 FT TURN 243 248
 FT STRAND 247 248
 FT STRAND 252 252
 FT STRAND 254 255
 FT STRAND 256 256
 FT STRAND 266 266
 FT TURN 283 283
 FT STRAND 284 286
 FT TURN 288 289
 FT STRAND 292 294
 FT TURN 297 299
 FT HELIX 300 309
 FT TURN 310 310
 FT TURN 312 313
 FT HELIX 317 320
 FT HELIX 323 332
 FT TURN 333 335
 FT TURN 337 338
 FT TURN 340 341
 FT TURN 347 349
 FT TURN 351 353
 FT HELIX 354 356
 FT HELIX 358 375
 FT TURN 376 377
 FT HELIX 383 386
 FT TURN 389 390
 FT STRAND 391 399
 FT TURN 403 404
 FT STRAND 405 416
 FT TURN 418 420
 FT TURN 424 425
 FT STRAND 429 438
 FT STRAND 442 449
 FT STRAND 455 465
 FT STRAND 467 467
 FT TURN 469 470
 FT STRAND 473 473
 FT HELIX 476 482
 FT STRAND 484 493
 FT STRAND 496 503
 FT HELIX 504 506
 FT TURN 508 509
 FT STRAND 510 511
 FT HELIX 517 521
 FT TURN 522 523
 FT STRAND 538 539
 FT HELIX 540 542
 FT STRAND 548 548
 FT TURN 549 550
 FT STRAND 552 562
 FT HELIX 563 566
 FT TURN 578 579
 FT HELIX 580 583
 FT TURN 586 587
 FT TURN 596 599
 FT HELIX 609 611
 FT TURN 612 612
 FT TURN 615 616
 FT STRAND 617 626
 SQ SEQUENCE 628 AA; 72628 MW; 6B8B4C6D8B1225BE CRC64;

Query Match

62.5%; Score 40; DB 1; Length 628;

Best Local Similarity 57.1%; Pred. No. 65;
Matches 4; Conservative 2; Mismatches

1; Indels 0; Gaps 0;

Qy 3 HKWHWYV 9

Db 172 HHWHHL 178

Search completed: December 16, 2003, 14:15:15
Job time : 5.66667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:09:45 ; Search time 25 seconds
(without alignments)
92.899 Million cell updates/sec

Title: US-09-870-089B-9
Perfect score: 64
Sequence: 1 FLHKWHYV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea.*
2: SP bacteria.*
3: SP fungi.*
4: SP human.*
5: SP invertebrate.*
6: SP mammal.*
7: SP mhc.*
8: SP organelle.*
9: SP phage.*
10: SP plant.*
11: SP rodent.*
12: SP virus.*
13: SP vertebrate.*
14: SP unclassified.*
15: SP rvirus.*
16: SP bacteriap.*
17: SP archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	71.9	328	10 Q9ZR87	Q9ZR87 zinnia eleg
2	45	70.3	542	11 Q9JQ0	Q9JQ0 mus musculus
3	44	68.8	147	2 Q8RP01	Q8RP01 bacillus an
4	44	68.8	159	16 Q9RS12	Q9RS12 deinococcus
5	44	68.8	498	10 Q9L5Q5	Q9L5Q5 arabidopsis
6	44	68.8	548	10 Q94A15	Q94A15 arabidopsis
7	44	68.8	554	4 Q8VWV7	Q8VWV7 homo sapien
8	44	68.8	554	4 Q92521	Q92521 homo sapien
9	44	68.8	561	5 Q9VZM5	Q9VZM5 drosophila
10	44	68.8	670	5 Q9BLD9	Q9BLD9 drosophila
11	44	68.8	681	5 Q77002	Q77002 hypanthia
12	44	68.8	683	5 Q9SR43	Q9SR43 drosophila
13	44	68.8	683	5 Q9W1V6	Q9W1V6 drosophila
14	44	68.8	683	5 Q964D5	Q964D5 galliera me
15	44	68.8	683	5 Q76951	Q76951 anopheles g
16	44	68.8	683	5 Q811F6	Q811F6 drosophila

17	44	68.8	684	5	Q96752	Q96752 anopheles g
18	44	68.8	684	5	Q9V521	Q9V521 drosophila
19	44	68.8	684	5	Q9GYW1	Q9GYW1 aedes aegypt
20	44	68.8	684	5	Q97047	Q97047 tenebrio mo
21	44	68.8	684	5	Q96751	Q96751 anopheles g
22	44	68.8	684	5	Q9N2P8	Q9N2P8 armigeres s
23	44	68.8	684	5	Q816K2	Q816K2 holotrichia
24	44	68.8	684	5	Q816K1	Q816K1 holotrichia
25	44	68.8	685	5	Q44249	Q44249 manduca sex
26	44	68.8	685	5	Q8MZM2	Q8MZM2 anopheles g
27	44	68.8	685	5	Q9GYW2	Q9GYW2 aedes aegypt
28	44	68.8	685	5	Q9C089	Q9C089 bombyx mori
29	44	68.8	685	5	Q9Y0B4	Q9Y0B4 sarcophaga
30	44	68.8	685	5	Q27451	Q27451 bombyx mori
31	44	68.8	686	5	Q8WSA8	Q8WSA8 aedes aegypt
32	44	68.8	686	5	Q44251	Q44251 anopheles g
33	44	68.8	686	5	Q96453	Q96453 anopheles g
34	44	68.8	687	5	Q96753	Q96753 anopheles g
35	44	68.8	695	5	Q25519	Q25519 manduca sex
36	44	68.8	696	5	Q8MZM4	Q8MZM4 anopheles g
37	44	68.8	697	5	Q76208	Q76208 hypanthia
38	44	68.8	699	5	Q9GVA5	Q9GVA5 pimpla hypo
39	44	68.8	700	5	Q8MZM3	Q8MZM3 anopheles g
40	44	68.8	708	5	Q9GVA7	Q9GVA7 pimpla hypo
41	43	67.2	146	2	Q8RPQ2	Q8RPQ2 bacillus an
42	43	67.2	147	16	Q9NSC0	Q9NSC0 staphylococ
43	43	67.2	148	16	Q8CNH5	Q8CNH5 staphylococ
44	43	67.2	209	16	Q9A3Z1	Q9A3Z1 caulobacter
45	43	67.2	213	16	Q8PIF4	Q8PIF4 xanthomonas

ALIGNMENTS

RESULT 1

Q9ZR87 ID Q9ZR87 PRELIMINARY; PRT; 328 AA.
AC Q9ZR87;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Bifunctional nuclease.
GN NUC2E2.
OS Zinnia elegans.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Zinnia.
OX NCBI_TaxID:34245;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20098724; PubMed=10631260;
RA Perez-Amador M.A., Abler M.L., De Rocher E.J., Thompson D.M.,
RA Van Hoof A., LeBrasseur N.D., Lers A., Green P.J.;
RT "Identification of BFN1, a bifunctional nuclease induced during leaf
RT and stem senescence in arabidopsis.";
RL plant physiol. 122:169-179(2000).
DR EMBL; U90266; AAC00695.1; -;
DR HSSP; P24289; IAKO.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR003154; S1/Flnuclease.
DR Pfam; PF02265; Nuclease; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
SQ SEQUENCE 328 AA; 37560 MW; 9343388936C53F7E CRC64;

Query Match 71.9%; Score 46; DB 10; Length 328;
Best Local Similarity 71.4%; Pred. No. 21;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLHKWHW 7

Db 75 WMEKWHW 81

```

RESULT 2
Q9JUQ0 PRELIMINARY; PRT; 542 AA.
AC Q9JUQ0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Pig-b.
GN PIGB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97015126; PubMed=8861954;
RA Takahashi M., Inoue N., Ohishi K., Maeda Y., Nakamura N., Endo Y.,
RA Fujita T., Takeda J., Kinoshita T.;
RT "PIG-B, a membrane protein of the endoplasmic reticulum with a large
RT luminal domain, is involved in transferring the third mannose of the
RT GPI anchor.";
RL EMBO J. 15:4254-4261(1996).
DR EMBL; D84436; BAA94827.1; -.
DR MGD; MGI:1891825; PigB.
DR InterPro; IPR005599; PMP.
DR Pfam; PF03901; PMP; 1.
SQ SEQUENCE 542 AA; 63133 MW; 0FC8BEB1D8A687C1 CRC64;

Query Match 70.3%; Score 45; DB 11; Length 542;
Best Local Similarity 71.4%; Pred. No. 48;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 LHKWHWYV 9
DB 295 HPWHWYL 301

RESULT 3
Q8RFQ1 PRELIMINARY; PRT; 147 AA.
AC Q8RFQ1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Maniferitin Dip2.
GN DIP2.
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21964043; PubMed=11836250;
RA Papinutto E., Dundon W.G., Fitulis N., Battistutta R., Montecucco C.,
RA Zanotti G.;
RT "Structure of Two Iron-binding Proteins from Bacillus anthracis.";
RL J. Biol. Chem. 277:15093-15098 (2002).
DR EMBL; AF374269; AAM18636.1; -.
DR InterPro; IPR002177; DPS.
DR InterPro; IPR001519; Ferritin.
DR Pfam; PF00210; ferritin; 1.
DR ProDom; PD149803; DPS; 1.
DR PROSITE; PS00819; DPS_1; 1.
SQ SEQUENCE 147 AA; 16649 MW; 2741651884FCCCD CRC64;

Query Match 68.8%; Score 44; DB 2; Length 147;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWYV 9
DB 25 LKNYHWYV 32

```

```

RESULT 4
Q9RS12 PRELIMINARY; PRT; 159 AA.
AC Q9RS12;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein DR2142.
GN DR2142.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RI;
RC MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Mofatt K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.W.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE002048; AAF11686.1; -.
DR TIGR; DR2142; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 159 AA; 17309 MW; FE0B9C87ED008F02 CRC64;

Query Match 68.8%; Score 44; DB 16; Length 159;
Best Local Similarity 62.5%; Pred. No. 21;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHWYV 9
DB 11 VHPWHWYV 18

RESULT 5
Q9LEQ5 PRELIMINARY; PRT; 498 AA.
AC Q9LEQ5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Dolichyl-phosphate-mannose-glycolipid alpha-mannosyltransferase-like
DE protein.
GN T9L3150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391149; CAC01884.1; -.
DR InterPro; IPR005599; PMP.
DR Pfam; PF03901; PMP; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 498 AA; 56575 MW; 55283E86D7238316 CRC64;

Query Match 68.8%; Score 44; DB 10; Length 498;

```

```

Best Local Similarity 83.3%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWY 8
DB 265 HPWHWY 270

RESULT 6
Q94A15 PRELIMINARY; PRT; 548 AA.
AC Q94A15;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE AT5g14850/T9L3.150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Koesema E., Chen H., Cheuk R., Kim C.J., Meyers M.C., Shinn P.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pam P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RA "Arabidopsis cDNA clones";
RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Banh J., Bowser L.,
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Pam P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY050456; AAK91470.1; -.
DR EMBL; AY120693; AAM52236.1; -.
DR InterPro; IPR005599; PMP.
DR Pfam; PF03901; PMP; 1.
SQ SEQUENCE 548 AA; 62701 MW; 125D3BEA0BF6A098 CRC64;

Query Match 68.8%; Score 44; DB 10; Length 548;
Best Local Similarity 83.3%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWY 8
DB 268 HPWHWY 293

RESULT 7
Q8WVW7 PRELIMINARY; PRT; 554 AA.
AC Q8WVW7;
DT 01-WAR-2002 (TReMBLrel. 20, Created)
DT 01-WAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Phosphatidylinositol glycan, class B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017711; AAH17711.1; -.
DR Genew; HGNC:8959; PIGH.
DR InterPro; IPR005599; PMP.
DR Pfam; PF03901; PMP; 1.
SQ SEQUENCE 554 AA; 64957 MW; E778418C02A27488 CRC64;

Query Match 68.8%; Score 44; DB 4; Length 554;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWY 8
DB 306 HPWHWY 311

RESULT 8
Q92521 PRELIMINARY; PRT; 554 AA.
AC Q92521;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE FIG-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-97015126; PubMed=8861954;
RA Takahashi M., Inoue N., Ohishi K., Maeda Y., Nakamura N., Endo Y.,
RA Fujita T., Takeda J., Kinoshita T.;
RT "FIG-B, a membrane protein of the endoplasmic reticulum with a large
RT lumenal domain, is involved in transferring the third mannose of the
RT GPI anchor.";
RL EMBO J. 15:4254-4261(1996).
DR EMBL; D42138; BAA07709.1; -.
DR InterPro; IPR005599; PMP.
DR Pfam; PF03901; PMP; 1.
SQ SEQUENCE 554 AA; 65056 MW; B2AF87D13ADF90B3 CRC64;

Query Match 68.8%; Score 44; DB 4; Length 554;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 HKWHWY 8
DB 306 HPWHWY 311

RESULT 9
Q9VZM5 PRELIMINARY; PRT; 561 AA.
AC Q9VZM5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE CG12006 protein.
GN CG12006.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-Berkeley;
RX MEDLINE-20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

```

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 [2]
 SEQUENCE FROM N.A.
 Celnikier S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 Bazan J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
 Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 McIntosh T.C., Moy M., Murphy B., Park S., Patel S., Pfeiffer B.,
 Pacleb J., Paragov V., Park S., Patel S., Pfeiffer B.,
 Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 "Sequencing of *Drosophila melanogaster* genome.";
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 SEQUENCE FROM N.A.
 Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 Hraderick P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikier S.E.,
 Ciamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 "Annotation of *Drosophila melanogaster* genome.";
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE FROM N.A.
 Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 SEQUENCE FROM N.A.
 FlyBase;
 Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 EMBL; AE003479; AAF47795.2; -.

DR FlyBase; FBgn0035464; CG12006.
 DR InterPro; IPR005599; PMP.
 SQ Pfam; PF03901; PMP; 1.
 Query Match 68.8%; Score 44; DB 5; Length 561;
 Best Local Similarity 83.3%; Pred. No. 70;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 HKWHWY 8
 DB 240 HPWHWY 245
 RESULT 10
 Q9BLD9 PRELIMINARY; PRT; 670 AA.
 AC Q9BLD9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Prophenol oxidase A3 (Fragment).
 GN DOX-A3 OR DOX-3 OR CG2952
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 SEQUENCE FROM N.A.
 RA Asada N., Hatta T., Norioka S., Kawamoto N.;
 RT "Properties activation and PCR-based sequence of prophenol oxidase A3
 in *Drosophila melanogaster*.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB055857; BAB43866.1; -.
 DR HSP; P04253; 1LLA.
 DR FlyBase; FBgn000487; Dox-A3.
 DR InterPro; IPR000896; Hemocyanin.
 DR InterPro; IPR005203; Hemocyanin_C.
 DR InterPro; IPR005204; Hemocyanin_N.
 DR InterPro; IPR002227; Tyrosinase.
 DR Pfam; PF03723; hemocyanin; 1.
 DR Pfam; PF03722; hemocyanin_C; 1.
 DR PRINTS; PR00187; HAEMOCYANIN.
 DR PROSITE; PS00209; HEMOCYANIN_1; 1.
 DR PROSITE; PS00210; HEMOCYANIN_2; 1.
 DR PROSITE; PS00498; TYROSINASE_2; 1.
 FT NON_TER 1
 FT NON_TER 670
 SQ SEQUENCE 670 AA; 77493 MW; FFA45D5CD4EBE69C8 CRC64;
 Query Match 68.8%; Score 44; DB 5; Length 670;
 Best Local Similarity 62.5%; Pred. No. 82;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LHKWHWY 9
 DB 206 LHHWHWHL 213
 RESULT 11
 Q77002 PRELIMINARY; PRT; 681 AA.
 AC Q77002;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Prophenoloxidase.
 GN HCPROPO.
 OS *Hyphantria cunea* (Fall webworm).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Noctuoidea;

```

OC Arctiidae; Hyphantria.
OX NCBI_TaxID=39466;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98199915; PubMed=9537761;
RA Park D.S., Shin S.W., Kim M.G., Park S.S., Lee W.J., Brey P.T.,
RA Park H.Y.;
RT "Isolation and characterization of the cDNA encoding the
RL phenoloxidase of fall webworm, hyphantria cunea.";
DR EMBL; U66875; AAC34251.1; -
DR HSP; P04253; 10XY.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
SQ SEQUENCE 681 AA; 78195 MW; D85A513318312E83 CRC64;

Query Match 68.8%; Score 44; DB 5; Length 681;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHYV 9
DB 207 LHHWHHL 214

RESULT 12
Q95R43 PRELIMINARY; PRT; 683 AA.
ID Q95R43
AC Q95R43;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE SD09866P.
GN DOX-A3 OR CG2952.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod W.F., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spidgen A.C., Stapleton M., Strong R., Sun E.,
RA Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
DR EMBL; AE003459; AAF46946.1; -.
DR HSP; P04253; 10XY.
DR FlyBase; FBgn0000487; Dox-A3.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
SQ SEQUENCE 683 AA; 79251 MW; E0E3F7E32540F20B CRC64;

Query Match 68.8%; Score 44; DB 5; Length 683;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHYV 9
DB 207 LHHWHHL 214

RESULT 13
Q9W1V6 PRELIMINARY; PRT; 683 AA.
ID Q9W1V6
AC Q9W1V6;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE CG2952 protein.
GN DOX-A3 OR CG2952.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod W.F., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spidgen A.C., Stapleton M., Strong R., Sun E.,
RA Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
DR EMBL; AE003459; AAF46946.1; -.
DR HSP; P04253; 10XY.
DR FlyBase; FBgn0000487; Dox-A3.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR Pfam; PF03722; hemocyanin_N; 1.
SQ SEQUENCE 681 AA; 78195 MW; D85A513318312E83 CRC64;

Query Match 68.8%; Score 44; DB 5; Length 681;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LHKWHYV 9
DB 207 LHHWHHL 214

```

```

DR PRINTS: PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
SQ SEQUENCE 683 AA; 79314 MW; 437CDD9E8A278BF CRC64;

Query Match      68.8%; Score 44; DB 5; Length 683;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWYV 9
Db 207 LHHWHWHL 214

RESULT 14
Q964D5 PRELIMINARY; PRT; 683 AA.
AC Q964D5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Prophenoloxidase.
OS Galleria mellonella (Wax moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyraloidea;
OC Pyralidae; Galleriinae; Galleria.
OX NCBI_TaxID=7137;
RN [1]
RP SEQUENCE FROM N.A.
RA Li D.M., Schmidt O., Theopold U.;
RT "Insect hemocyte-derived microparticles are regulatory assemblages of
RL the cell-derived immune response.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF336289; AAK64363.1; -.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
SQ SEQUENCE 683 AA; 78545 MW; 54A1F80F992839B7 CRC64;

Query Match      68.8%; Score 44; DB 5; Length 683;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWYV 9
Db 207 LHHWHWHL 214

RESULT 15
O76951 PRELIMINARY; PRT; 683 AA.
AC O76951;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE PROPHENOLOXIDASE.
GN AGPROPO.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RA
RT
RL
SQ SEQUENCE FROM N.A.
RA Lee W.J., Ahmed A., della Torre A., Kobayashi A., Ashida M.,
RA Brey P.T.;
RT "Molecular cloning and chromosomal localization of a prophenoloxidase
RT cDNA from the malaria vector Anopheles gambiae.";
RL Insect Mol. Biol. 7:41-50(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Ahmed A., Lee W.-J., Brey P.T.;
RT "Genomic structure of a prophenoloxidase from the malaria vector,
RT Anopheles gambiae.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; I76038; AAC27383.1; -.
DR EMBL; AF031626; AAD01936.1; -.
DR HSSP; P04253; 10XY.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00372; hemocyanin; 1.
DR Pfam; PF03722; hemocyanin_C; 1.
DR PRINTS; PR00187; HAEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
SQ SEQUENCE 683 AA; 79330 MW; 59494CF4869AB7DD CRC64;

Query Match      68.8%; Score 44; DB 5; Length 683;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LHKWHWYV 9
Db 205 LHHWHWHL 212

Search completed: December 16, 2003, 14:18:01
Job time : 26 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:35 ; Search time 5.66667 Seconds
(without alignments)
74.689 Million cell updates/sec

Title: US-09-870-089B-13
Perfect score: 55
Sequence: 1 KHFKPHGFS 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	100.0	351	1	ATF4 HUMAN
2	49	89.1	349	1	ATF4 MOUSE
3	38	69.1	341	1	PUR7 VIGAC
4	37	67.3	110	1	SAA HORSE
5	37	67.3	642	1	MB1L ARATH
6	36	65.5	342	1	Y4NF RHISN
7	36	65.5	361	1	NAGZ NEIMA
8	36	65.5	361	1	NAGZ NEIMA
9	35	63.6	48	1	RL34 MYCPU
10	35	63.6	153	1	SODC NEUCR
11	35	63.6	244	1	YBGL ECOLI
12	35	63.6	264	1	RS31 ARATH
13	35	63.6	350	1	RS40 ARATH
14	35	63.6	368	1	VE2 HPV52
15	35	63.6	512	1	ENAL HUMAN
16	35	63.6	512	1	ENAL RAT
17	35	63.6	649	1	H570 PARBR
18	35	63.6	718	1	LSP2 DROME
19	35	63.6	868	1	MCE ASFB7
20	34	61.8	45	1	RL34 STAM
21	34	61.8	211	1	YCSF BACSU
22	34	61.8	297	1	ACTR CAVPO
23	34	61.8	354	1	PHOH ECOLI
24	34	61.8	379	1	YXK0 YEAST
25	34	61.8	521	1	SPKA SYN3
26	34	61.8	953	1	SVV VIBCH
27	34	61.8	1463	1	PA2R BOVIN
28	34	61.8	2410	1	POLL BAYWJ
29	33	60.0	92	1	ASH BACSU
30	33	60.0	262	1	LAMB EMENI
31	33	60.0	289	1	UBIG RICCN
32	33	60.0	299	1	H5F6 ARATH
33	33	60.0	352	1	GAL7 CAEBL

RESULT 1
ATF4_HUMAN
ID ATF4_HUMAN STANDARD; PRT; 351 AA.
AC P18848; Q9UH31;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cyclic-AMP-dependent transcription factor ATF-4 (Activating transcription factor 4) (DNA-binding protein TAXREB67) (Cyclic AMP response element-binding protein 2) (CREB2).
GN ATF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91240735; PubMed=1847461;
RA Tsujimoto A., Niyunoya H., Morita T., Sato T., Shimotohno K.;
RT "Isolation of cDNAs for DNA-binding proteins which specifically bind to a tax-responsive enhancer element in the long terminal repeat of human T-cell leukemia virus type 1.";
RL J. Virol. 65:1420-1426(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92279218; PubMed=1534408;
RA Karpinski B.A., Morle G.D., Huggenvik J., Uhler M.D., Leiden J.M.;
RT "Molecular cloning of human CREB-2: an ATF/CREB transcription factor that can negatively regulate transcription from the CAMP response element.";
RL Prof. Natl. Acad. Sci. U.S.A. 89:4820-4824 (1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dykham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M., Clapp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K., Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P., Bird C.P., Blakey S.E., Bridgman M., Buck D., Burgess J., Furrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R., Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G., Evans K.L., Fey J.M., Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C., Hall R.E., Hall-Ramlyn G., Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A., Laird G.K., Langford C.F., Levertha M.A., Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T., McClay J., McLaren S., McMurray A., Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T., Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L., Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L., Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,

ALIGNMENTS

Q39473 cinamomum
Q9hfs2 cochliobolu
P55495 rhizobium s
Q9wzr9 thermotoga
Q92gl1 rickettsia
P51692 homo sapien
P42232 mus musculu
P52632 rattus norv
Q9tuz0 bos taurus
Q9tuz0 sus scrofa
Q13797 homo sapien
Q9h9y6 homo sapien

34 33 60.0 382 1 PATB CINCA
35 33 60.0 430 1 CREA COCCA
36 33 60.0 432 1 Y4IL RHISN
37 33 60.0 451 1 UXAC THEMA
38 33 60.0 670 1 SYGB RICCEN
39 33 60.0 786 1 ST5B HUMAN
40 33 60.0 786 1 ST5B MOUSE
41 33 60.0 786 1 ST5B RAT
42 33 60.0 787 1 ST5B BOVIN
43 33 60.0 787 1 ST5B PIG
44 33 60.0 1035 1 ITA9 HUMAN
45 33 60.0 1079 1 RPA2 HUMAN

```

DR PRINTS; PR00306; SERUMAMYLOID.
DR ProDom; PR002112; Serum_amyloid_A; 1.
DR SMARI; SM00197; SAA; 1.
DR PROSITE; PS00992; SAA; 1.
DR Acute Phase; Plasma; HDL; Amyloid.
FT CHAIN 1 110 SERUM AMYLOID A PROTEIN.
FT VARIANT 16 16 I -> L.
FT VARIANT 44 44 K -> Q.
FT VARIANT 59 59 G -> G.
FT VARIANT 78 78 G -> A.
SQ SEQUENCE 110 AA; 12289 MW; EAE7DBE7AB007ESB CRC64;

Query Match 67.3%; Score 37; DB 1; Length 110;
Best Local Similarity 83.3%; Pred. No. 4.8;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 HKFPHG 7
DB 100 HRFPHG 105

RESULT 5
MBIL_ARATH
ID MBIL_ARATH STANDARD; PRT; 642 AA.
AC Q9SAV1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myrosinase binding protein-like At1G52030.
GN AT1G52030 OR F5F19.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.D., Brooks S.Y.,
RA Buehler E., Chao L., Chen K., Chen R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408, 816-820(2000).
CC -!- SIMILARITY: BELONGS TO THE JACALIN LECTIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC006216; AAD12677.1;
CC FTR; A96560.
CC HSSP; P18674; 1JOT.
CC InterPro; IPR001229; Jacalin_lectin.

```

```

DR PRINTS; PR00306; SERUMAMYLOID.
DR ProDom; PR002112; Serum_amyloid_A; 1.
DR SMARI; SM00197; SAA; 1.
DR PROSITE; PS00992; SAA; 1.
DR Acute Phase; Plasma; HDL; Amyloid.
FT CHAIN 1 110 SERUM AMYLOID A PROTEIN.
FT VARIANT 16 16 I -> L.
FT VARIANT 44 44 K -> Q.
FT VARIANT 59 59 G -> G.
FT VARIANT 78 78 G -> A.
SQ SEQUENCE 110 AA; 12289 MW; EAE7DBE7AB007ESB CRC64;

Query Match 67.3%; Score 37; DB 1; Length 110;
Best Local Similarity 83.3%; Pred. No. 4.8;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 HKFPHG 7
DB 100 HRFPHG 105

RESULT 5
MBIL_ARATH
ID MBIL_ARATH STANDARD; PRT; 642 AA.
AC Q9SAV1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myrosinase binding protein-like At1G52030.
GN AT1G52030 OR F5F19.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.D., Brooks S.Y.,
RA Buehler E., Chao L., Chen K., Chen R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408, 816-820(2000).
CC -!- SIMILARITY: BELONGS TO THE JACALIN LECTIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC006216; AAD12677.1;
CC FTR; A96560.
CC HSSP; P18674; 1JOT.
CC InterPro; IPR001229; Jacalin_lectin.

```

```

DR Pfam; PF01419; Jacalin; 4.
DR LECTIN; Repeat; Multigene family.
DR SEQUENCE 642 AA; 68849 MW; 9AB0B569A65C913C CRC64;

Query Match 67.3%; Score 37; DB 1; Length 642;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HKFPHG 7
DB 157 HKFPHG 163

RESULT 6
Y4WF_RHISN
ID Y4WF_RHISN STANDARD; PRT; 342 AA.
AC P55684;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 37.7 kDa protein Y4WF.
GN Y4WF.
OS Rhizobium sp. (strain NGR234).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387, 394-401(1997).
CC -!- SIMILARITY: TO ALKANAL MONOOXYGENASE ALPHA AND BETA CHAINS
(CC 1.14.14.3) (BACTERIAL LUCIFERASE). BUT DISTANTLY RELATED.
CC -!- SIMILARITY: TO Y4WJ.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000103; AAB91913.1;
CC InterPro; IPR002103; Bac_luciferase.
CC Pfam; PF00296; bac_luciferase; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Plasmid.
SQ SEQUENCE 342 AA; 37665 MW; EB1D57A71EDA89C3 CRC64;

Query Match 65.5%; Score 36; DB 1; Length 342;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 HKFPHGFS 9
DB 52 HFSPHGFS 59

RESULT 7
NAGZ_NEIMA
ID NAGZ_NEIMA STANDARD; PRT; 361 AA.
AC Q9JVT3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-hexosaminidase (EC 3.2.1.52) (N-acetyl-beta-glucosaminidase)
GN NAGZ OR NMA0708
OS Neisseria meningitidis (serogroup A)
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

```


CC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 Davies R.M., Davis P., Devlin K., Feitwell T., Hamlin M., Holroyd S.,
 Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 Whitehead S., Spratt B.G., Barrall B.G.;
 RA "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis Z2491";
 RT Nature 404:502-506(2000).
 RL Nature 404:502-506(2000).
 CC -!- FUNCTION: Cleaves GlcNAc linked beta-1,4 to Murnac tripeptides
 (By similarity).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-
 acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides.
 CC -!- PATHWAY: Cell wall synthesis; murein tripeptide recycling pathway.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF GLYCOSYL HYDROLASES. NAGZ
 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AL162754; CAB83994.1; -
 DR PIR; H81913; H81913.
 DR HAMAP; MF_00364; -; 1.
 DR InterPro; IPR001764; Glyco_hydro_3N.
 DR Pfam; PF00933; Glyco_hydro_3; 1.
 DR PROSITE; PS00775; GLYCOSYL_HYDROL_F3; FALSE_NEG.
 KW Hydrolase; Glycosidase; Peptidoglycan synthesis; Cell division;
 KW Cell wall; Complete proteome.
 FT ACT SITE 258 258 BY SIMILARITY.
 SQ SEQUENCE 361 AA; 39022 MW; 174F048B59CAAFBC CRC64;
 Query Match 65.5%; Score 36; DB 1; Length 361;
 Best Local Similarity 75.0%; Pred. No. 23;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KHFKPHGF 8
 DB 174 KHFKPHGF 181
 RESULT 8
 NAGZ NEIMB
 ID NAGZ NEIMB STANDARD; PRT; 361 AA.
 AC Q9K0Q4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Beta-hexosaminidase (EC 3.2.1.52) (N-acetyl-beta-glucosaminidase)
 DE (Beta-N-acetylhexosaminidase).
 GN NAGZ OR NMB0530.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OC NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,

RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
 Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
 Gill J., Scarlato V., Masiagnani V., Pizzo M., Grandi G., Sun L.,
 Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58";
 RL Science 287:1809-1815(2000).
 CC -!- FUNCTION: Cleaves GlcNAc linked beta-1,4 to Murnac tripeptides
 (By similarity).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-
 acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides.
 CC -!- PATHWAY: Cell wall synthesis; murein tripeptide recycling pathway.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF GLYCOSYL HYDROLASES. NAGZ
 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE002408; AAF40960.1; -
 DR PIR; B81190; B81190.
 DR TIGR; NMB0530; -
 DR HAMAP; MF_00364; -; 1.
 DR InterPro; IPR001764; Glyco_hydro_3N.
 DR Pfam; PF00933; Glyco_hydro_3; 1.
 DR PROSITE; PS00775; GLYCOSYL_HYDROL_F3; FALSE_NEG.
 KW Hydrolase; Glycosidase; Peptidoglycan synthesis; Cell division;
 KW Cell wall; Complete proteome.
 FT ACT SITE 258 258 BY SIMILARITY.
 SQ SEQUENCE 361 AA; 39152 MW; 3E066E5B9809033A CRC64;
 Query Match 65.5%; Score 36; DB 1; Length 361;
 Best Local Similarity 75.0%; Pred. No. 23;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KHFKPHGF 8
 DB 174 KHFKPHGF 181
 RESULT 9
 RL34 MYCPU
 ID RL34 MYCPU STANDARD; PRT; 48 AA.
 AC Q98R56;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L34.
 GN RPMH OR MYPU 1540.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OC NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CT1P;
 RX MEDLINE=21267165; PubMed=11353084;
 RA Chandaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
 Mosser I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
 Blanchard A.;
 RA "The complete genome sequence of the murine respiratory pathogen
 Mycoplasma pulmonis";
 RT Mycoplasma pulmonis; 23:2145-2153(2001).
 RL Nucleic Acids Res. 29:2145-2153(2001).
 CC -!- SIMILARITY: BELONGS TO THE L34P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AL445563; CAC13327.1; --
 DR PIR; B90531; B90531.
 DR MyPUList; MYPU 1540; --
 DR HAMAP; MF_00391; --; 1.
 DR InterPro; IPR000271; Ribosomal L34.
 DR Pfam; PF00468; Ribosomal L34; 1.
 DR ProDom; PD003101; Ribosomal L34; 1.
 DR TIGRFAMs; TIGR01030; rpmH Bact; 1.
 DR PROSITE; PS00784; RIBOSOMAL L34; 1.
 KW Ribosomal protein, Complete Proteome.
 SQ SEQUENCE 48 AA; 5581 MW; 72F420BD6EC4DD43 CRC64;

Query Match 63.6%; Score 35; DB 1; Length 48;
 Best Local Similarity 75.0%; Pred. No. 4.9;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HKFKPHGF 8

DB 12 KHAKTHGF 19

RESULT 10

SODC NEUCR STANDARD; PRT; 153 AA.
 AC P07509;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
 GN SOD-1.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91035418; PubMed=2146266;
 RA Chavry P., Hallewell R.A., Natvig D.O.;
 RT "Structure, exon pattern, and chromosome mapping of the gene for
 RT cytosolic copper-zinc superoxide dismutase (sod-1) from Neurospora
 RT crassa."; J. Biol. Chem. 265:18961-18967(1990).
 RL J. Biol. Chem. 265:18961-18967(1990).
 RN [2]
 RP SEQUENCE.
 RX MEDLINE=85261356; PubMed=3160699;
 RA Lerch K., Schenk E.;
 RT "Primary structure of copper-zinc superoxide dismutase from
 RT Neurospora crassa."; J. Biol. Chem. 260:9559-9566(1985).
 RL J. Biol. Chem. 260:9559-9566(1985).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M58687; AAAG3780.1; --

DR PIR; A36591; A36591.

DR HSSP; P00445; 1JCV
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; sodcu; 1.
 DR PRINTS; PR00068; CUZNDISMTASE.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN; 1.
 DR PROSITE; PS00332; SOD_CU_ZN; 2; 1.
 KW Antioxidant; Oxidoreductase; Metal-binding; Copper; Zinc.

FT INIT_MET 0
 FT METAL 46 46 COPPER (BY SIMILARITY).
 FT METAL 48 48 COPPER (BY SIMILARITY).
 FT METAL 63 63 COPPER AND ZINC (BY SIMILARITY).
 FT METAL 71 71 ZINC (BY SIMILARITY).
 FT METAL 80 80 ZINC (BY SIMILARITY).
 FT METAL 83 83 ZINC (BY SIMILARITY).
 FT METAL 120 120 COPPER (BY SIMILARITY).
 FT DISULFID 57 146 BY SIMILARITY.
 FT CONFLICT 69 69 T -> A (IN REF. 2).
 SQ SEQUENCE 153 AA; 15753 MW; 56DAFAC86CD9573B CRC64;

Query Match 63.6%; Score 35; DB 1; Length 153;
 Best Local Similarity 83.3%; Pred. No. 15;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 HKFKPHG 7

DB 63 HFNPHG 68

RESULT 11

YBGL ECOLI STANDARD; PRT; 244 AA.
 ID YBGL_ECOLI
 AC P75746;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein ybgl.
 GN YBGL OR B0713.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MGL655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."; Science 277:1453-1474 (1997).
 RL Science 277:1453-1474 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map."; DNA Res. 3:137-155(1996).
 RL DNA Res. 3:137-155(1996).
 RN [3]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RX MEDLINE=99420866; PubMed=10493123;
 RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
 RT "Enrichment of low abundance proteins of Escherichia coli by
 RT hydroxyapatite chromatography."; Electrophoresis 20:2181-2195(1999).
 RL Electrophoresis 20:2181-2195(1999).
 CC -!- SIMILARITY: STRONG, TO H.INFLUENZAE HI1729.

```

CC -!- SIMILARITY: TO E.NIDULANS LACTAM UTILIZATION PROTEIN LAMB AND
CC B.SUBTILIS YCSF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000174; AAC73807.1; -.
DR EMBL; D90710; BAA35377.1; -.
DR PIR; H64806; H64806.
DR EcGene; EGI3308; ybgl.
DR InterPro; IPR005501; Lamb_YcsF.
DR Pfam; PF03746; Lamb_YcsF; 1.
KW Complete proteome.
SQ SEQUENCE 244 AA; 25800 MW; DBC0FD14ADA0EB0A CRC64;
Query Match 63.6%; Score 35; DB 1; Length 244;
Best Local Similarity 71.4%; Pred. No. 24;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KHFKPHG 7
Db 101 RHVKPHG 107
RESULT 12
RS31 ARATH STANDARD; PRT; 264 AA.
AC P29264; Q944A4; Q9M278;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginine/serine-rich splicing factor RSP31.
DE RSP31 OR AT3G61860 OR F21F14.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97143875; PubMed=8989882;
RA Lopato S., Waigmann E., Barta A.;
RT "Characterization of a novel arginine/serine-rich splicing factor in
RT Arabidopsis.";
RL Plant Cell 8:2255-2264(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unsel M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choisme N., Artiguenave F., Robert C., Brotier P.,
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Warmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudle M., Berger-Blauro C., Fumelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Marnhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P.,

```

```

RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:820-822(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAPFs) sequenced by the
RT SSP consortium (Saik/Stanford/FGEC).";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR CONSTITUTIVE AND ALTERNATIVE PRE-MRNA
CC SPLICING.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: Highly expressed in roots and flowers. A
CC presumably longer alternatively spliced form is found in leaves,
CC stems and flowers.
CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC -!- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99435; CAA67798.1; -.
DR EMBL; ALI38642; CAB71893.1; ALT_SEQ.
DR EMBL; AF439831; AAL27502.1; -.
DR EMBL; AY125565; AAM78075.1; -.
DR PIR; T51304; T51304.
DR HSP; P09651; 1HA1.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; FALSE NEG.
DR Nuclear protein; RNA-binding; Alternative splicing; Repeat.
FT DOMAIN 2 74 RNA-BINDING (RRM) 1.
FT DOMAIN 93 164 RNA-BINDING (RRM) 2.
FT DOMAIN 172 264 ARG/SER-RICH (RS DOMAIN).
FT CONFLICT 66 66 R -> K (IN REF. 1).
SQ SEQUENCE 264 AA; 31154 MW; 445FD2B4A5E0C203 CRC64;
Query Match 63.6%; Score 35; DB 1; Length 264;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KHFKPHG 7
Db 112 KHFEYFG 118
RESULT 13
RS40 ARATH STANDARD; PRT; 350 AA.
ID RS40 ARATH
AC P29265; Q96333;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginine/serine-rich splicing factor RSP40.

```

GN RSP40 OR AT4G25500 OR WTJ2.130.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=971143875; PubMed=8989882;
RA Lopato S., Waigmann B., Barta A.;
RT "Characterization of a novel arginine/serine-rich splicing factor in
Arabidopsis.";
RL Plant Cell 8:2255-2264 (1996).
RN [2]
RP REVISIONS.
RC STRAIN=cv. Columbia;
RA Barta A.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambert R., Murphy G., Volckaert G.,
RA Pohl T., Duisterhoef A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoerg W., Brandt P., Grivell L.A., Rieger M., Mueller M.,
RA Weichselgartner M., de Simone V., Obermayer B., Mache R., Schmidtchnei T.,
RA Kreis M., Delseny S., Puigdomenech P., Watson M., Schmidchnei T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Vost M., Bastiaens I., Aert R., Defoor E.,
RA Meitzenger T., Bothe G., Rampsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirksse W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Partmann B., Granderath K., Dauner D., Heitzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefor P., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Perez-Perez A., Fumelle B., Bent B., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Stacker S.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Flocks C.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sakon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latrelle P., Courtney L., cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton D., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana.";

RL Nature 402:769-777 (1999).
CC -!- FUNCTION: REQUIRED FOR CONSTITUTIVE AND ALTERNATIVE PRE-MRNA
CC SPLICING (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: Highly expressed in roots and flowers. A
CC presumably longer alternatively spliced form is found in leaves,
CC stems and flowers.
CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC -!- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X99437; CAA67800.1; -;
CC EMBL; U76607; AAB18813.1; -;
CC EMBL; AL022197; CAA18176.1; -;
CC PIR; T05797; T05797.
CC InterPro: IPR000504; RNA_rec_mot.
CC Pfam: PF00076; rrm; 2.
CC SMART; SM00360; RRM; 2.
CC DR PROSITE; PS00102; RRM; 2.
CC DR PROSITE; PS00030; RRM RNP 1; FALSE NEG.
CC KW Nuclear protein; RNA-binding; Alternative splicing; Repeat.
CC DOMAIN 2 74 RNA-BINDING (RRM) 1.
CC DOMAIN 97 168 POLY-SER.
CC FT DOMAIN 89 92 ARG/SER-RICH (RS DOMAIN).
CC FT DOMAIN 173 350 K -> R (IN REF. 3).
CC FT CONFLICT 27 27 K -> R (IN REF. 3).
CC FT CONFLICT 63 63 R -> T (IN REF. 3).
CC FT CONFLICT 66 66 R -> T (IN REF. 3).
CC FT CONFLICT 195 195 S -> T (IN REF. 3).
CC SQ SEQUENCE 350 AA; 40319 MW; 4A1C779BB94B32A6 CRC64;
Query Match 63.6%; Score 35; DB 1; Length 350;
Best Local Similarity 71.4%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KHPKPHG 7
Db 116 KHPKPHG 122
RESULT 14
VE2 HPV52
ID_VPV52 STANDARD; PRT; 368 AA.
AC P36736;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 52.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OC NCBI_TaxID=10618;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94265501; PubMed=8205838;
RA Delius H., Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RL Curr. Top. Microbiol. Immunol. 186:13-31 (1994).
CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 14:06:35 ; Search time 5.66667 Seconds
(without alignments)
74.689 Million cell updates/sec

Title: US-09-870-089B-13

Perfect score: 55

Sequence: 1 KHFKPHGFS 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	55	100.0	351	1	ATF4_HUMAN	P18848	homo sapien
2	49	89.1	349	1	ATF4_MOUSE	Q06507	mus musculus
3	38	69.1	341	1	PUR7_VIGAC	Q07463	vigna con
4	37	67.3	110	1	SAA_HORSE	P19857	equus caball
5	37	67.3	642	1	MB11_ARATH	Q9SAV1	arabidopsis
6	36	65.5	342	1	Y4WF_RHISN	P55684	rhizobium s
7	36	65.5	361	1	NAGZ_NEIMA	Q9JVE3	neisseria m
8	36	65.5	361	1	NAGZ_NEIMA	Q9K094	neisseria m
9	35	63.6	48	1	RL34_MYCPU	Q98R56	mycoplasma
10	35	63.6	153	1	SODC_NEUCR	P07509	neurospora
11	35	63.6	244	1	YBGL_ECOLI	P75746	escherichia
12	35	63.6	264	1	R831_ARATH	P92964	arabidopsis
13	35	63.6	350	1	R840_ARATH	P92965	arabidopsis
14	35	63.6	368	1	V22_HPV52	P36796	human papill
15	35	63.6	512	1	BNAI_HUMAN	O16515	homo sapien
16	35	63.6	512	1	BNAI_HUMAN	Q62962	rattus norv
17	35	63.6	649	1	H570_PARRR	P87047	rattus norv
18	35	63.6	718	1	LSP2_DROME	Q24388	paracoccidi
19	35	63.6	868	1	MCE_ASFB7	P32094	african swi
20	34	61.8	45	1	RL34_STAMU	Q99GT1	staphylococ
21	34	61.8	211	1	YCSF_BACSU	P42963	bacillus su
22	34	61.8	297	1	ACTR_CAVPO	Q92149	cavia porce
23	34	61.8	354	1	PHXO_ECOLI	P31544	escherichia
24	34	61.8	379	1	YKNO_YEAST	P3861	saccharomyc
25	34	61.8	521	1	SPKA_SYNT3	Q9FAB3	synthecocyst
26	34	61.8	953	1	SVY_VIBCH	Q9K773	vibrio chol
27	34	61.8	1463	1	PA2R_BOVIN	P49259	bos taurus
28	34	61.8	2410	1	POL1_BAYMJ	Q01206	barley yell
29	33	60.0	92	1	ABH_BACSU	P39758	bacillus su
30	33	60.0	262	1	LAME_EMENI	P38096	emeritella
31	33	60.0	289	1	UBIG_RICCN	Q9ZHO7	rickettsia
32	33	60.0	299	1	HSF6_ARATH	Q9SCW4	arabidopsis
33	33	60.0	352	1	GAL7_CAEEL	Q27536	caenorhabdi

RESULT 1
ATF4_HUMAN
ID ATF4_HUMAN STANDARD; PRT; 351 AA.
AC P18848; O9UH31;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cyclic-AMP-dependent transcription factor ATF-4 (Activating
transcription factor 4) (DNA-binding protein TAXREB67) (Cyclic AMP
response element-binding protein 2) (CREB2).
GN ATF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140735; PubMed=1847461;
RA Tsujimoto A., Niyunoya H., Morita T., Sato T., Shimotohno K.;
RT "Isolation of cDNAs for DNA-binding proteins which specifically bind
to a tax-responsive enhancer element in the long terminal repeat of
human T-cell leukemia virus type I.";
RL J. Virol. 65:1420-1426(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92279218; PubMed=1534408;
RA Karpinski B.A., Morle G.D., Huguenik J., Uhler M.D., Leiden J.M.;
RT "Molecular cloning of human CREB-2: an ATF/CREB transcription factor
that can negatively regulate transcription from the cAMP response
element.";
RL Prof. Natl. Acad. Sci. U.S.A. 89:4820-4824(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Duhham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
CJamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
Bisguoy C., Bailey J., Barlow K.P., Bates K.N., Beasley O.P.,
Bird C.P., Blakey S.E., Bridgeman M.M., Buck D., Burgess J.,
Burkhill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
Clegg S.M., Cobley V.B., Cole C.G., Collier R.E., Connor R.,
Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
Gilbert J.G.R., Goward M.E., Grahnam D.V., Griffiths M.N.D., Hall C.,
Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
Hunt S.E., Jones M.C., Kersey J., Kimberley A.M., King A.,
Laird G.E., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,

ALIGNMENTS

34	33	60.0	382	1	FATB_CINCA	Q39473	cinnamomum
35	33	60.0	430	1	CREA_COCCA	Q9HFS2	cochliobolu
36	33	60.0	432	1	Y4IL_RHISN	P55495	rhizobium s
37	33	60.0	451	1	UXAC_THEMA	Q9WXR9	thermotoga
38	33	60.0	670	1	SYGB_RICCN	Q92G11	rickettsia
39	33	60.0	786	1	STSB_HUMAN	P51692	homo sapien
40	33	60.0	786	1	STSB_MOUSE	P42232	mus musculus
41	33	60.0	786	1	STSB_RAT	P52632	rattus norv
42	33	60.0	787	1	STSB_BOVIN	Q9TUM3	bos taurus
43	33	60.0	787	1	STSB_PIG	Q9TUZO	sus scrofa
44	33	60.0	1035	1	ITA9_HUMAN	Q13797	homo sapien
45	33	60.0	1079	1	RPA2_HUMAN	Q9H5Y6	homo sapien

RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulten R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA Korfi I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA McErmid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.S.,
RA Wilkison P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [4]
RN SEQUENCE FROM N.A.
RP TISSUE=Lung, and Placenta;
RX MEDLINE=22386257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Sheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mollath S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Halle S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RN SEQUENCE OF 274-341 FROM N.A.
RX MEDLINE=90185187; PubMed=2516827;
RA Hai T., Liu F., Coukos W.J., Green M.R.;
RT "Transcription factor ATF cDNA clones: an extensive family of leucine
zipper proteins able to selectively form DNA-binding heterodimers.";
RL Genes Dev. 3:2083-2090(1989).
RN [6]
RN ERRATUM.
RA Hai T., Liu F., Coukos W.J., Green M.R.;
RL Genes Dev. 4:682-682(1990).
CC -!- FUNCTION: THIS PROTEIN BINDS TO A TAX-RESPONSIVE ENHANCER
CC ELEMENT IN THE LONG TERMINAL REPEAT OF HUMAN T-CELL LEUKEMIA
CC VIRUS TYPE 1.
CC -!- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
CC (CONSENSUS: 5'-GTGAGT(A/C)(A/G)-3'), A SEQUENCE PRESENT IN MANY
CC VIRAL AND CELLULAR PROMOTERS.
CC -!- SUBUNIT: Binds DNA as a dimer.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the bZIP family.

CC EMBL; D90209; BAA14234.1; -
DR EMBL; M86842; AAA52071.1; -
DR EMBL; AL022312; CAB45284.1; -
DR EMBL; BC080930; AAH08090.1; -
DR EMBL; BC011994; AAH11994.1; -
DR EMBL; BC024775; AAH24775.1; -
DR EIR; A45377; A45377.
DR PDB; 1CI6; 04-DEC-00.
DR TRANSFAC; T01303; -
DR Genew; HGNC:786; ATF4.
DR MIM; 604064; -
DR GO; GO:0003677; F:DNA binding activity; TAS.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; TAS.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bZIP; 1.
DR SMART; SM00338; bZIP; 1.
DR PROSITE; PS00217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Polymorphism; 3D-structure.
FT DNA_BIND 280 300 BASIC MOTIF.
FT DOMAIN 306 334 LEUCINE-ZIPPER (PROBABLE).
FT VARIANT 322 322 E -> D (in dbSNP:1803324).
FT /FTID=VAR_014768.
FT CONFLICT 22 22 P -> Q (IN REF. 3).
FT CONFLICT 284 284 K -> R (IN REF. 2).
FT CONFLICT 290 290 T -> R (IN REF. 5).
FT CONFLICT 329 331 KEI -> REK (IN REF. 5).
FT CONFLICT 338 338 I -> L (IN REF. 5).
SQ SEQUENCE 351 AA; 38558 MW; 7A708C5CD6ED7F8 CRC64;
Query Match 100.0%; Score 55; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KHFKPHGFS 9
DB 42 KHFKPHGFS 50
RESULT 2
ATF4_MOUSE STANDARD; PRT; 349 AA.
ID ATF4_MOUSE
AC Q66507; Q61906;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyclic-AMP-dependent transcription factor ATF-4 (C/EBP-related ATF)
DE (C/ATF) (TAXREB67 homolog).
GN ATF4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC TISSUE=Adipocyte;
RX MEDLINE=93281642; PubMed=8506317;
RA Vallejo M., Ron D., Miller C.P., Habener J.F.;
RT "c-ATF, a member of the activating transcription factor family of
RT DNA-binding proteins, dimerizes with C/AT/enhancer-binding proteins
RT and directs their binding to cAMP response elements.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:4679-4683(1993).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Embryo;
RX MEDLINE=92335183; PubMed=1631061;
RA Chevray P.M., Nathans D.;
RT "Protein interaction cloning in yeast: identification of mammalian
RT proteins that react with the leucine zipper of Jun.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:5789-5793(1992).
CC -!- FUNCTION: BINDS TO ASYMMETRIC CAMP RESPONSE ELEMENTS (CRE) AS